

**UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
MESTRADO EM BIOINFORMÁTICA**

LUIZ FERNANDO DA SILVA SOARES

**BIOINFORMÁTICA APLICADA À UNIFICAÇÃO DOS BANCOS DE
HISTOCOMPATIBILIDADE DOS LABORATÓRIOS DE IMUNOGENÉTICA DO
HOSPITAL DE CLÍNICAS E GENÉTICA MOLECULAR HUMANA DA UFPR**

CURITIBA

2011

LUIZ FERNANDO DA SILVA SOARES

**BIOINFORMÁTICA APLICADA À UNIFICAÇÃO DOS BANCOS DE
HISTOCOMPATIBILIDADE DOS LABORATÓRIOS DE IMUNOGENÉTICA DO
HOSPITAL DE CLÍNICAS E GENÉTICA MOLECULAR HUMANA DA UFPR**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós Graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, Universidade Federal do Paraná, área de concentração Bioinformática.

Orientadora: Dra. Jeroniza Nunes Marchaukoski

Co-Orientadora: Dra. Maria Berenice R. Steffens

CURITIBA

2011

S676 Soares, Luiz Fernando da Silva
Bioinformática aplicada à unificação dos bancos de histocompatibilidade dos laboratórios de Imunogenética do Hospital de Clínicas e Genética Molecular Humana da UFPR / Luiz Fernando da Silva Soares. - Curitiba, 2011. - 135 f.: il., tabs, grafs.

Orientador: Profa. Dra. Jeroniza Nunes Marchaukoski.
Co-orientador: Profa. Dra. Maria Berenice Reynaud Steffens.
Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Educação Profissional e Tecnológica, Curso de Pós-Graduação em Bioinformática.
Inclui Bibliografia.

1. Human Leykocyte Antigen. 2. HLA. 3. Unification. 4. Profile. 5. Antígenos de histocompatibilidade HLA . 6. Bioinformática.
I. Marchaukoski, Jeroniza Nunes. II. Steffens, Maria Berenice Reynaud. III. Título. IV. Universidade Federal do Paraná.

CDD 616.079


TERMO DE APROVAÇÃO

LUIZ FERNANDO DA SILVA SOARES


Bioinformática aplicada à unificação dos bancos de histocompatibilidade dos laboratórios de Imunogenética do Hospital de Clínicas e Genética Molecular Humana da UFPR

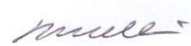
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Bioinformática, pelo Programa de Pós-graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador:


Prof.ª Dr.ª Jeroniza Nunes Marchaukoski

Coorientador:


Prof.ª Dr.ª Maria Berenice Reynaud Steffens


Prof. Dr. Roberto Tadeu Raittz
Universidade Federal do Paraná


Prof.ª Dr.ª Karin Braun Prado
Universidade Federal do Paraná


Prof.ª Dr.ª Lupe Furtado Alle
Universidade Federal do Paraná

Curitiba, 28 de fevereiro de 2011

Aos meus pais e amigos pelo apoio.
Dedico em especial a minha namorada
pela paciência e suporte durante
toda a vida acadêmica.

AGRADECIMENTOS

A minha orientadora que me auxiliou em todas as etapas desta pesquisa.

A minha co-orientadora sempre prestativa.

Aos laboratórios que forneceram os dados na pessoa das profs Dras Noemi Farah Pereira e Maria Luiza Petzl-Erler, bem como aos técnicos de laboratório Joselito Getz e Cleusa Schafhauser.

A secretaria do Programa de Pós Graduação em Bioinformática.

Ao programa REUNI, CNPQ e CAPES pelo auxílio financeiro durante a pesquisa.

Aos meus colegas do mestrado em Bioinformática pelo apoio e auxílio na pesquisa.

Aos meus familiares pelo apoio e carinho.

A minha namorada pelas dicas na formatação, sugestões e testes do sistema, bem como pelo carinho e paciência.

Se eu vi mais longe, foi por estar de pé sobre ombros de gigantes.

Isaac Newton.

RESUMO

O Laboratório de Imunogenética do Hospital das Clínicas (LIHC) da Universidade Federal do Paraná (UFPR) dispõe de um grande volume de dados de tipagens HLA de familiares de pacientes, que foram cadastrados em um banco com o objetivo de serem doadores de medula. Por esse motivo, é desejável que este banco de doadores teoricamente neutro e não viciado possa ser utilizado em estudos de genética de populações como banco de controle. O Laboratório de Genética Molecular Humana (LGMH) da UFPR possui um banco de dados de indivíduos do Paraná que é utilizado como controle. Este trabalho teve como propósito principal a unificação dos bancos de dados do LIHC com o banco de dados do LGMH, com o objetivo secundário de determinar se o banco do LIHC pode ser empregado como banco controle (grupo de indivíduos). Foram criadas páginas e scripts para unificação, classificação e limpeza dos dados. Foi utilizado banco de dados PostgreSQL e linguagem de programação PHP. Um perfil dos bancos foi traçado através do programa PyPop e Arlequin. Descobriu-se que o banco do LIHC pode ser usado como controle em virtude da análise realizada. Porém a análise demonstrou que houve diferença estatística somente para três especificidades distintas das 43 observada para a etnia branca e nenhuma para a etnia mulata. Este resultado foi obtido para o cálculo de frequências comparativas do genes HLA-DRB1 em comparativo com o banco controle do LGMH. Por fim, os scripts e páginas criados para separação e organização dos dados permitem a realização de pesquisas em tempo menor, bem como a garantia de dados padronizados no futuro.

Palavras-chaves: HLA unificação perfil Antígenos de histocompatibilidade HLA

ABSTRACT

The Laboratório de Imunogenética do Hospital das Clínicas (LIHC) from Universidade Federal do Paraná (UFPR) has a huge volume of data of HLA from relatives of patients, that were registered in a database with the goal to be donors of bone marrow and others. Therefore, it is expected that this donor bank theoretically neutral and unbiased can be used in studies of genetics population as a control database. The Laboratório de Genética Molecular Humana (LGMH) of UFPR has a database with data from individuals of Paraná, which is used as a control group. This work aimed to unify the databases LIHC with the LGMH database, with the secondary purpose to determine if LIHC database can be used as a control database (group of individuals). Pages and scripts were created for the unification, classification and data cleaning. We used the PostgreSQL as the database and PHP as the programming language. A profile of the banks was traced through the program PyPop and Arlequin. It was discovered that the bank of LIHC can be used as a control database with the analysis done. However, the analysis showed that there were only three statistically distinct specificities observed for the 43 observed in the Caucasian ethnicity and none for any mulatto. This result was obtained with frequency comparative calculation of antigen DRB1 from LGMH control database. Finally, the scripts and pages created for sorting and organizing data allows the development of researches take less time, and ensuring standardized data in the future.

Key-words: *HLA unification profile histocompatibility antigens HLA*

SUMÁRIO

AGRADECIMENTOS.....	6
RESUMO	8
ABSTRACT.....	9
SUMÁRIO	10
LISTA DE ILUSTRAÇÕES.....	12
LISTA DE TABELAS.....	13
LISTA DE ABREVIATURAS E SIGLAS	15
1 INTRODUÇÃO.....	15
1.1 COMPLEXO PRINCIPAL DE HISTOCOMPATIBILIDADE	15
1.2 FREQUÊNCIA ALÉLICA E FREQUÊNCIA HAPLOTÍPICA	19
1.3 BANCO DE DADOS CONTROLE.....	19
1.4 DELIMITAÇÃO DO TEMA E FORMULAÇÃO DO PROBLEMA	20
1.5 OBJETIVOS.....	20
1.5.1 Objetivo Geral	21
1.5.2 Objetivos específicos	21
2 JUSTIFICATIVA.....	23
3 MATERIAIS E MÉTODOS	25
3.1 DADOS.....	25
3.2 POPULAÇÃO DE ESTUDO.....	26
3.2.1 Propriedades das informações e confidencialidade.....	29
3.3 MATERIAIS UTILIZADOS	29
3.3.1 Do Hardware.....	29
3.3.2 Do Software	30
3.3.2.1 Sistema operacional Windows 7 Profissional	30
3.3.2.2 Microsoft Office 2007 Enterprise.....	30

3.3.2.2.1 Access 2007	31
3.3.2.2.2 Word 2007	31
3.3.2.2.3 Excel 2007	31
3.3.2.3 ARLEQUIN 3.5.1.2.....	31
3.3.2.4 PYPOP 0.7.0.....	32
3.3.2.5 POSTGRESQL 9	32
3.3.2.6 PHP 5.2	36
3.3.2.7 WAMP 2.0i.....	37
3.3.2.8 API GOOGLE V 3.0 e WebService dos Correios.....	38
3.4 ESTATÍSTICA.....	39
3.5 A CRIAÇÃO DO BANCO UNIFICADO	41
4 RESULTADOS	45
4.1 FLUXO DE DADOS - O SISTEMA CRIADO E O BANCO DE DADOS	45
4.1.1 Gerador de .Arp	49
4.1.2 Plotagem de dados	55
4.1.3 Exportação de dados para o Excel	60
4.2 ANÁLISES ESTATÍSTICAS.....	62
4.3 BANCO DVMO DO LIHC.....	62
4.4 BANCO DA IMUNOGENÉTICA DO LIHC (TOTAL DE INDIVÍDUOS SELECIONADOS – UM DE CADA FAMÍLIA OU OS PAIS QUANDO NÃO CONSANGUÍNEOS)	72
4.5 BANCO DA IMUNOGENÉTICA DO HC (FAMILIARES DE PACIENTES SEM AAS E AF)	76
4.6 BANCO DO LGMH	91
4.7 ANÁLISE DE COMPARAÇÃO ENTRE LIHC E LGMH	96
4.8 ANÁLISE PACIENTES DE PÊNFIGO, LEUCEMIA E ANEMIA DE FANCONI.....	97
4.8.1 Pênfigo.....	97

4.8.2 Leucemia e Anemia de Fanconi	97
4.9 ANÁLISE DE GEOPOSICIONAMENTO	102
5 DISCUSSÃO	103
5.1 FREQUENCIA ALÉLICA.....	103
5.2 FREQUENCIA HAPLOTÍPICA.....	106
5.3 PÁGINAS DO SISTEMA.....	107
5.4 BANCO DE CONTROLE	108
6 CONCLUSÃO	109
BLIBLIOGRAFIA.....	110
APÊNCICE	118
APÊNDICE 1 – DECLARAÇÃO ORIENTADOR	119
APÊNDICE 2 – MODELO CARTA ENCAMINHAMENTO.....	120
APÊNDICE 3 – MODELO DE DISPENSA DE CONSENTIMENTO.....	121
APÊNDICE 4 – MODELO IDENTIFICAÇÃO DE PROJETO.....	122
APÊNDICE 5 – FOLHA DE ROSTO COMITÊ	124
APÊNDICE 6 - HAPLÓTIPO SIMPLES	125
APÊNDICE 7 - HAPLÓTIPO DUPLO	126
APÊNDICE 8 - FREQUÊNCIA HLA-A PÊNFIGO	127
APÊNDICE 9 - FREQUÊNCIA HLA-B PÊNFIGO	128
APÊNDICE 10 - FREQUÊNCIA HLA-A LEUCEMIA E ANEMIA DE FANCONI	129
APÊNDICE 11 - FREQUÊNCIA HLA-B LEUCEMIA E ANEMIA DE FANCONI	130
APÊNDICE 12 - TABELA COMPARATIVA DRB1 BRANCO LIHC e LGMH	132
APÊNDICE 13 - TABELA COMPARATIVA DRB1 MULATO LIHC e LGMH.	134

LISTA DE ILUSTRAÇÕES

FIGURA 1 - Mapa MHC	16
FIGURA 2 - Quantidade de indivíduos distribuído por origem dos dados	26
FIGURA 3. Gráfico comparativo etnias LGMH banco controle	28
FIGURA 4 - Simplificação da unificação	33
FIGURA 5 - Fluxo de trabalho	42
FIGURA 6 - Fluxo de dados	46
FIGURA 7 - Passos realizados no desenvolvimento	47
FIGURA 8 - Caso de uso páginas desenvolvidas	48
FIGURA 9 - Caso de uso exportar dados	48
FIGURA 10 - Diagrama Entidade Relacionamento	49
FIGURA 11 - Passos de criação de arquivo ".arp"	50
FIGURA 12 - Descrição arquivo ".arp"	52
FIGURA 13 - Específica 1 .arp	53
FIGURA 14 - Específica 2 .arp	54
FIGURA 15 - Dupla .arp	54
FIGURA 16 - Gerar arquivo de haplótipo.....	55
FIGURA 17 - Obtenção de coordenadas geográficas	57
FIGURA 18 - Configuração plotagem	58
FIGURA 19 - Plotagem Simples	59
FIGURA 20 - Plotagem dupla loci.....	60
FIGURA 21 - Criação de arquivo de planilha eletrônica	61
FIGURA 22 - Dados exportados.....	62
FIGURA 23 - Comparativo de trabalhos	104
FIGURA 24 - Haplótipo simples.....	125
FIGURA 25 - Haplótipo duplo	126
 QUADRO 1. Quantidade de indivíduos agrupados por etnia e origem.....	27
QUADRO 2. Quadro comparativo de resultados obtidos no banco	36
QUADRO 3. Tabela Qui-Quadrado	40
QUADRO 4. Descrição cabeçalho arquivo ".Arp"	51

LISTA DE TABELAS

TABELA 1. Frequência HLA-A DVMO	62
TABELA 2. Frequência HLA-B DVMO	63
TABELA 3. Frequência HLA-DRB1 DVMO	65
TABELA 4. Frequência haplotípica HLA-A/B/DRB1	65
TABELA 5. Frequência HLA-A - DVMO por etnias	67
TABELA 6. Frequência HLA-B DVMO por etnias	69
TABELA 7. Frequência HLA-DRB1 DVMO por etnias	71
TABELA 8. Frequência HLA-A Amostra 2.....	72
TABELA 9. Frequência HLA-B Amostra 2.....	73
TABELA 10. Frequência HLA-DRB1 Amostra 2	75
TABELA 11. Frequência haplotípica HLA-A*B*DRB1 Amostra 2.....	76
TABELA 12. Frequência HLA-A Amostra 3.....	76
TABELA 13. Frequência HLA-B Amostra 3.....	78
TABELA 14. Frequência HLA-DRB1 Amostra 3	79
TABELA 15. Frequência haplotípica HLA-A*B*DRB1 Amostra 3.....	81
TABELA 16. Frequência HLA-A Amostra 3 por etnia.....	84
TABELA 17. Frequência HLA-B Amostra 3 por etnia.....	87
TABELA 18. Frequência HLA-DRB1 Amostra 3 por etnia.....	89
TABELA 19. Frequência HLA-DRB1 LGMH	91
TABELA 20. Frequência HLA-DRB1 LGMH por etnias.....	93
TABELA 21. Frequência HLA-DR LGMH por etnia.....	95
TABELA 22. Frequência HLA-A e HLA-B Leucemia e Fanconi comparativa.	97
TABELA 23. Frequência DRB1 Pênfigo Leucemia e Anemia de Fanconi.....	99
TABELA 24. Comparativo Antígenos Goldberg, Monte e DVMO.....	105
TABELA 25. Frequência haplotípica HLA-A;B;DRB1 Leucemia e Anemia Fanconi	107
TABELA 26. Frequência HLA-A Pênfigo.....	127
TABELA 27. Frequência HLA-B Pênfigo.....	128
TABELA 28. Frequência HLA-A Leucemia e Anemia de Fanconi.....	129
TABELA 29. Frequência HLA-B Leucemia e Anemia de Fanconi.....	130
TABELA 30. Tabela comparativa DRB1 etnia branca - LIHC e LGMH	132

TABELA 31. Tabeka comparativa DRB1 etnia mulato - LIHC e LGMH134

LISTA DE ABREVIATURAS E SIGLAS

AAS - ANEMIA APLÁSTICA SEVERA

AF - ANEMIA DE FANCONI

CEP - CÓDIGO DE ENDEREÇAMENTO POSTAL

CNPq - CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO

CAPES - COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR

DEPTO. - DEPARTAMENTO

DVMO - DOADOR VOLUNTÁRIO DE MEDULA ÓSSEA

HC - HOSPITAL DAS CLÍNICAS

HLA - *HUMAN LEUKOCYTE ANTIGEN*

HW - HARDY WEINBERG

LGMH - LABORATÓRIO DE GENÉTICA MOLECULAR HUMANA

LIHC - LABORATÓRIO DE IMUNOGENÉTICA DO HOSPITAL DAS CLÍNICAS

LLA - LEUCEMIA LINFÓIDE OU LINFOBLÁSTICA AGUDA

LMA - LEUCEMIA MIELÓIDE AGUDA

LMC - LEUCEMIA MIELÓIDE CRÔNICA

MHC - MAJOR HISTOCOMPATIBILITY COMPLEX

REDOME - REGISTRO NACIONAL DE DOADORES DE MEDULA ÓSSEA

SBT - *SEQUENCE BASED TYPING*

SIG - SISTEMA DE INFORMAÇÃO GEOGRÁFICOS

SSOP - *SEQUENCE SPECIFIC OLIGONUCLEOTIDE PROBE TECHNIQUE*

SSP-PCR - *SEQUENCE SPECIFIC PRIMERS/POLYMERASE CHAIN REACTION*

SQL - *STRUCTURED QUERY LANGUAGE*

TMO - TRANSPLANTE DE MEDULA ÓSSEA

UFPR - UNIVERSIDADE FEDERAL DO PARANÁ

1 INTRODUÇÃO

Este trabalho é resultado de um estudo abordando o sistema de antígenos leucocitários humanos (HLA), do inglês *Human Leukocyte Antigen*. Trata-se de um estudo elaborado envolvendo três centros. Bancos de dados de HLA provenientes do Laboratório de Imunogenética do Hospital das Clínicas (LIHC) e do Laboratório de Genética Molecular Humana (LGMH) que foram comparados e estudados pelo Programa de Pós Graduação em Bioinformática da Universidade Federal do Paraná. Um trabalho mais voltado a unificação dos bancos do que a uma análise propriamente dita. Este trabalho está dividido em Introdução, onde serão abordados os conceitos fundamentais para entendimento da pesquisa, Materiais e Métodos, onde serão descritos os materiais e os métodos empregados, seguido por Resultados, Discussão e Conclusão.

1.1 COMPLEXO PRINCIPAL DE HISTOCOMPATIBILIDADE

Os genes HLA estão localizados no complexo MHC(Major Histocompatibility Complex) ou em português no Complexo Principal de Histocompatibilidade. O MHC localiza-se no braço curto do cromossomo 6 (6p21.3) (BICALHO *et al.*, 2002) e é responsável pela codificação das moléculas de apresentação de antígenos ao sistema imune (ALVES, 2006).

O MHC compreende mais de 200 genes (TSUNETO *et al.*, 2003) sendo que os considerados clássicos são os que tem maior variabilidade. Os genes podem ser específicos de uma região ou população ou ainda estarem distribuídos amplamente em uma dada população (TSUNETO *et al.*, 2003).

Os genes HLA estão agrupados em três regiões: Classe I, Classe II e Classe III. Sendo a classe I composta pelos genes HLA-A, HLA-B e HLA-C, a classe II é composta pelos genes HLA-DM, HLA-DO, HLA-PD, HLA-DQ e HLA-DR. No presente estudo serão abordados principalmente os genes A, B e DRB1. A classe I e II são referentes aos genes envolvidos com a resposta imune (DONADI, 2000). A **FIGURA 1** apresenta o mapa do complexo de MHC, onde é possível observar, nitidamente, as 3 classes existentes.

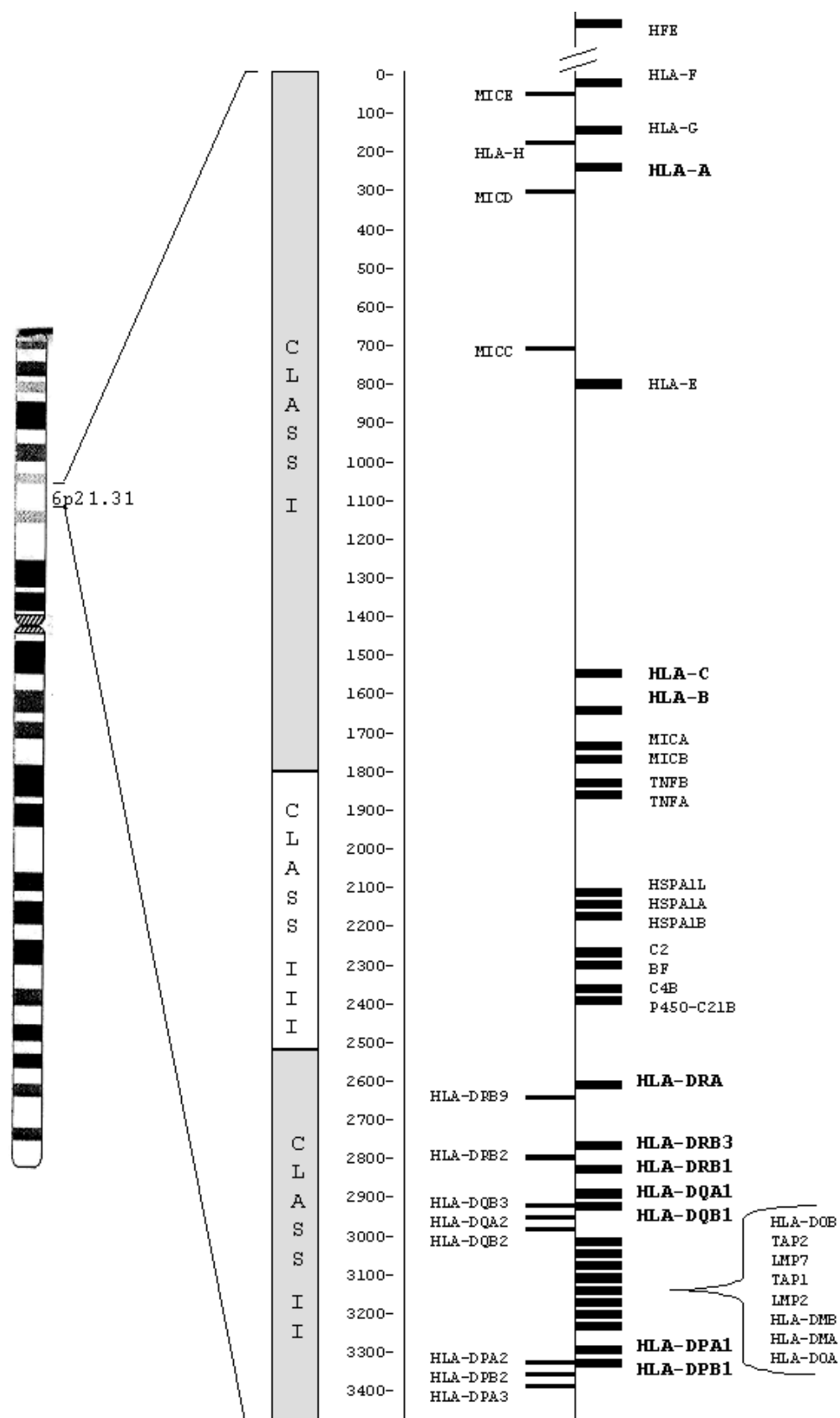


FIGURA 1 - Mapa MHC

Fonte: http://www.questdiagnostics.com/brand/business/hla_immu_gen/reg_map.html

As principais moléculas de histocompatibilidade envolvidas no transplante são as que são codificadas pelos genes pertencentes aos locos A, B e C da classe I e da classe II pelas regiões DR e DQ (SELL *et al.*, 2006). Algumas delas foram estudadas nesse trabalho.

No transplante de medula óssea (TMO) a compatibilidade do HLA é verificada por sorologia ou biologia molecular, através da coleta do sangue periférico. Os genes HLA estão presentes na superfície de todas as células nucleadas e são herdadas do pai e da mãe. Um dos fatores que justifica o grande número de indivíduos de familiares de pacientes do Hospital das Clínicas (n=35871) é a necessidade dos pacientes encontrarem um familiar com compatibilidade mínima para realização de um transplante. No Brasil, a dificuldade de se encontrar um doador compatível é decorrente da grande diversidade étnica existente, além da baixa quantidade de indivíduos que se cadastram no banco nacional de doadores voluntários de medula o REDOME (RUIZ *et al*, 2005). Além disso, ainda há a mistura em cruzamentos étnicos internos que por sua vez criam novas combinações haplotípicas. A grande diversidade étnica resultante teve como origem principal basicamente três grupos: brancos; indígenas e negros, porém a intensa miscigenação originou diversos outros grupos.

Para realização de um transplante de medula o doador deve estar vivo, diferentemente no caso de um transplante de órgão. Além disso, deverá haver entre o doador e o paciente uma total compatibilidade do HLA, isso para que doenças do tipo auto-imune não se manifestem. Uma doença comum na realização de um transplante é a Doença do Enxerto-Contra-Hospedeiro (DECH), a maior causa de morbidade de pacientes que realizaram transplante de medula (SANTOS *et al*, 2005). É caracterizada pelo ataques às células do hospedeiro pelas células do doador, pois estas reconhecem-no como corpo estranho. O efeito inverso também ocorre, e por isso imunodepressores são aplicados ao hospedeiro para que o transplante ocorra com uma maior taxa de sucesso.

Nesse estudo uma das doenças co-relacionadas é a leucemia. Um dos tratamentos para ela é o TMO, por isso um dos fatores geradores de volume de dados no banco. A leucemia trata-se de uma neoplasia maligna no sangue. Existem basicamente 3 tipos de leucemia, a Linfóide ou Linfoblástica Aguda(LLA), a Mielóide Aguda (LMA) e a Mielóide Crônica(LMC).

Também envolvida nesse estudo está a doença anemia aplástica severa (AAS), trata-se de uma doença com características tais como insuficiência de glóbulos vermelhos, brancos e plaquetas. Difere de uma leucemia, pois esta última é caracterizada pela produção de células anormais e não pela carência tal como a anemia. É originada por fármacos específicos, exposição a químicos tóxicos, radiação e doenças do tipo auto-imune. Pode ser congênita ou adquirida, no caso a probabilidade de ser congênita de 20% e adquirida 80% (GUEDES *et al*, 2008). Este grupo é composto por uma amostragem de 993 indivíduos, cerca de 2,31% do banco de dados do HC.

Uma das doenças considerada nesse estudo é a anemia de Fanconi (AF), trata-se de uma doença autossômica recessiva ou ligada ao cromossomo X. É caracterizada pela falência progressiva da medula óssea, por malformação congênita e predisposição a neoplasias, como leucemia mielóide aguda. Também tem como característica um fenótipo confuso, para confirmação desta doença são necessários testes citogenéticos, um dos fatores para se ter uma grande concentração destes dados no banco de estudo. Esta anemia é estudada pelo LIHC e é composta por (n=399) indivíduos.

Na quantidade de (n= 360) estão os pacientes de pênfigo, esse banco é originário do LGMH. Pênfigo é uma doença onde se manifestam diversas bolhas de variados tamanhos sobre a pele, mucosa da boca e outras membranas. Essa doença é causada por um ataque auto-imune contra as estruturas superficiais da epiderme.

Um número grande (n=1066) de pessoas forma o grupo de estudo Pesquisa B-27. Esses dados são de origem do LIHC. O antígeno HLA-B27 está frequentemente associado a algumas doenças, tais como espondilite anquilosante, síndrome de Reiter, artrite reumatóide e auto-imune. Nos últimos anos esse antígeno vem sendo bastante estudado (ALVAREZ *et al*, 2000; BALL *et al*, 2001; BLANCO-GELAZ *et al*, 2001; CHEN *et al*, 2002; SAMPAIO-BARROS *et al*, 2003 e TAMOUZA *et al*, 2001).

A tipagem do HLA, como dito anteriormente, pode ocorrer de duas maneiras, sendo elas o método sorológico e o método molecular. O método sorológico tem como característica a baixa resolução. Por sua vez o método molecular pode ser dividido em Sequence Specific Primers/Polymerase Chain Reaction (SSP-PCR), Sequence Specific Oligonucleotide Probe Technique (SSOP) e Sequence Based

Typing (SBT). Tanto o método sorológico como o método molecular foram empregados na obtenção dos dados que constituem esse trabalho.

1.2 FREQUÊNCIA ALÉLICA E FREQUÊNCIA HAPLOTÍPICA

Frequência alélica é a contagem da frequência que um dado alelo de um certo locus aparece em uma população. Já a frequência haplotípica trata-se de inferir a frequência em que uma combinação de alelos aparece em dada população analisando diferentes locus e alelos.

Inferir os haplótipos é muito importante para diversas suposições genéticas, inclusive para associação de fenótipos e regiões gênicas (ITO *et al.*, 2003). É muito importante para análises de desequilíbrio de ligação (ZHANG *et al.*, 2003). O desequilíbrio de ligação ocorre quando um genótipo de um locus é dependente do genótipo de outro locus. Isso acarreta que a frequência de combinação dos alelos é diferente da frequência esperada.

Além disso, o cálculo das frequências alélicas e haplotípicas pode ser usado, por exemplo, para informar um paciente da probabilidade de se encontrar um doador compatível dado um grupo étnico específico ou para traçar metas e estratégias de recrutamento de doadores (MORI *et al.*, 1997). Também é interessante para estudo da seleção e evolução de genes polimórficos e associação com doenças (MORI *et al.*, 1997). Os cálculos utilizados e as fórmulas específicas para contagem dos alelos serão descritas na sessão de materiais e métodos sub-capítulo 3.4 ESTATÍSTICA.

1.3 BANCO DE DADOS CONTROLE

Geralmente, pesquisas que envolve teste de medicamentos ou análises populacionais referentes a doenças ou relação de algum fator de impacto sobre uma população, apresentam um grupo chamado de grupo controle. Para o presente estudo um grupo de controle, trata-se de uma amostra de indivíduos saudáveis. É uma amostra não viciada e teoricamente neutra que pode ser utilizada para comparação com outras populações, por exemplo comparação entre indivíduos saudáveis com indivíduos doentes.

1.4 DELIMITAÇÃO DO TEMA E FORMULAÇÃO DO PROBLEMA

O presente trabalho teve por objetivo principal a unificação das três bases de dados. Como consequência também permitir detectar, através da associação entre a tipagem HLA e os grupos étnicos existentes, uma relação de similaridade. Essa análise foi realizada a partir do banco de dados de doadores voluntários de medula óssea (DVMO), pacientes e familiares de pacientes do Laboratório de Imunogenética do Hospital de Clínicas (LIHC) da Universidade Federal do Paraná (UFPR), situado à Rua General Carneiro, 181 - Curitiba – PR. Esta análise foi confrontada com o banco de dados do Laboratório de Genética Molecular Humana (LGMH) da mesma Universidade, situado à Rua Coronel Francisco Hoffmann dos Santos, 210. Estudos semelhantes já foram apresentados por diversos pesquisadores (PRADO *et al.*, 1999; PROBST *et al.*, 2000; BOLDT *et al.*, 2009; DALAIO *et al.*, 2002; ALVES, 2006; FERREIRA *et al.*, 2006 e BRAUN-PRADO *et al.*, 2000) muitos deles realizados com dados de populações do Paraná, e em alguns casos encontrando-se diversas divergências (PRADO *et al.*, 1999 e DALAIO *et al.*, 2002). O principal diferencial nesse estudo, em relação as análises subsequentes após a unificação dos bancos é o confronto de dois bancos de dados distintos, aliados a quantidade expressiva de indivíduos do banco de dados do LIHC, cerca de 32000, indivíduos pacientes e familiares de pacientes do HC e outros 2500 indivíduos referentes a doadores voluntários de medula óssea.

Com o objetivo principal pretende-se facilitar a análise para se traçar um padrão de antígenos dos leucócitos humanos com o banco de dados a ser estudado, e depois correlacioná-los com os estudos já realizados por demais pesquisadores referentes a determinadas regiões como Paraná, Rio de Janeiro com os genes DRB1, DQA1, DQB1 e DPB1 (MORAES *et al.*, 1993), Piauí para os genes HLA-A, B, DRB1 e DQB1 (MONTE *et al.*, 2004) São Paulo e Rondônia com os genes DRB1, DQA1 e DQB1 (GOLDBERG *et al.*, 1998) e demais encontrados na literatura. Esse estudo tem uma alta importância para o TMO, a medida em que permite conhecermos melhor o perfil da população brasileira.

1.5 OBJETIVOS

Abaixo serão descritos os objetivos deste trabalho, divididos em objetivo geral e objetivos específicos.

1.5.1 Objetivo Geral

Unificação dos bancos de dados dos bancos de dados do LIHC, DVMO e LGMH. Com consequencia de determinar as frequências alélicas dos genes HLA-A, HLA-B, HLA-C, HLA-DR, HLA-DRB1, HLA-DQ, HLA-DQA1, HLA-DQB1 e associá-las aos grupos étnicos para traçar um perfil dos bancos de dados.

1.5.2 Objetivos específicos

Amostra 1: Banco Doadores Voluntários de Medula Óssea (DVMO) fornecido pelo LIHC

- Determinar frequências de grupos alélicos dos genes HLA-A, B e DRB1;
- Determinar frequências de haplótipos constituídos pelos grupos alélicos dos genes HLA-A, B e DRB1.

Amostra 2: Banco LIHC (total de indivíduos selecionados – um de cada família ou os pais quando não consanguíneos)

- Determinar frequências de grupos alélicos dos genes HLA-A, B e DRB1;
- Determinar frequências de haplótipos constituídos pelos grupos alélicos dos genes HLA-A, B e DRB1;

Amostra 3: Banco LIHC (familiares de pacientes sem AAS e AF)

- Determinar frequências de grupos alélicos dos genes HLA-A, B e DRB1;
- Determinar frequências de haplótipos constituídos pelos grupos alélicos dos genes HLA-A, B e DRB1.

Amostra 4: Banco do LGMH – Depto. Genética

- Determinar frequências de grupos alélicos do gene DRB1.

Banco de Dados LIHC, LGMH, DVMO:

Associação com localização dos grupos e das tipagens obtidas, além de posterior confirmação de associação entre algumas doenças e algumas tipagens HLA encontradas na literatura. Isso através da unificação dos três bancos com padronização dos dados e eliminação de redundâncias.

Sistema de Geoposicionamento:

O principal objetivo neste caso é a priorização da conversão do endereço do indivíduo em uma coordenada geográfica com posterior criação de um sistema de plotagem dos dados de tipagem HLA na API Google Maps. Por meio da qual o pesquisador possa encontrar padrões de geo posicionamento das amostras a serem estudadas.

Utilização do banco de dados de familiares de pacientes do LIHC como banco de controle:

A partir das análise haplotípicas determinar se é possível utilizar o banco de dados de possíveis doadores de pacientes do LIHC como banco de dados de controle para futuras pesquisas.

2 JUSTIFICATIVA

O presente trabalho apresenta a correta divisão e a associação dos grupos étnicos às tipificações HLA, confirmando os diversos estudos encontrados na literatura. Também apresenta a resposta a uma das questões de pesquisa de que o banco de dados do LIHC seja um modelo de banco de referência que possa ser utilizado como banco controle, no caso os dados referentes aos familiares de pacientes, bem como confirmação de outras associações que também são encontradas na literatura entre grupo étnico e tipagem e a associação a algumas doenças.

A pesquisa é justificada em virtude da necessidade do LIHC ter um panorama do banco de dados de pacientes, conhecer como os antígenos leucocitários humanos encontram-se distribuídos por meio do cálculo das frequências alélicas e haplotípicas. Ou seja, a necessidade de se ter o banco contabilizado e com as frequências gênicas dos genes HLA clássicos. O presente gerou diversas análises estatísticas possibilitará uma maior quantidade de trabalhos e o consequente uso das pesquisas e dos dados observados como um banco de controle ou ainda para inferência de outras possíveis observações. A análise do banco permitirá que no futuro comparações de frequência HLA possam novamente ser realizadas com outros bancos, ou com o mesmo banco, porém em tempo de coleta distintos. Ainda, reconhecimento de possíveis padrões na população que possam vir a permitir estudos mais aprofundados e a melhoria das técnicas de buscas de doadores.

Estudos semelhantes de associação entre grupos étnicos e tipagem HLA vem sendo desenvolvidos no mundo todo a exemplo México (VARGAS-ALARCÓN *et al.*, 2006; BARQUERA *et al.*, 2008; INFANTE *et al.*, 1999 e VARGAS-ALARCÓN *et al.*, 2003) , Japão (NAITO, 2005), Líbano (FAYZA *et al.*, 2005), África (SANCHEZ-MAZAS, 2001 e PARRA *et al.*, 1998) , China (MCMANUS *et al.*, 2001), Argentina (PAROLIN, 2009), Paraguai (PAROLIN, 2009), Peru (PABLO *et al.*, 2000), Jordânia (SANCHEZ-VELASCO *et al.*, 2001), Venezuela (FERNÁNDEZ-MESTRE *et al.*, 2004), entre outros. Tais estudos permitem um panorama melhor da população, é possível através desses estudos entender processo de seleção natural, entender melhor a nossa história, como por exemplo, qual a real etnia de um indivíduo entre

outros. Com este estudo é possível conhecer de que forma é composta a população de uma dada região, bem como sua origem.

Tais objetivos só são alcançados graças a unificação destes bancos com consequência da padronização e organização dos dados.

3 MATERIAIS E MÉTODOS

Nos subcapítulos que se seguem serão descritos os materiais e métodos empregados na realização deste trabalho.

3.1 DADOS

Os dados englobam tipagens HLA, grupo étnico, sexo e CEP (Código de Endereçamento Postal) em alguns casos.

Sendo:

Banco da Imunogenética do HC

Pacientes (n=6706) e possíveis doadores familiares (n=32469), dados contém:

- Tipagem: HLA-A,B,C,DR,DRB1,DQB1;
- Sexo;
- Etnia;
- Doença;
- Parentesco;
- Coleta a partir de 2006 até dias atuais.

Banco DVMO

Doadores voluntários de medula óssea (n=2418), dados contém:

- Código Doador;
- Tipagem: HLA-A,B,DRB1
- Sexo;
- Etnia;
- CEP;
- Coleta sem informação de data.

Banco LGMH

Banco controle de indivíduos saudáveis (n=603) e pacientes de pênfigo (n=651):

- Código Indivíduo
- Tipagem: HLA-A,B,C,DR,DRB1,DQ,DQA1,DQB1.
- Sexo;

- Etnia;
- Nascimento
- Coleta sem informação de data.

A distribuição dos dados conforme origem é demonstrada na **FIGURA 2**. Apesar do grande número de indivíduos de familiares de pacientes do HC, nem todos encontram-se tipados em todos os genes. Isso ocorre em virtude de logo no início já ser detectado que há incompatibilidade entre doador e possível doador. Por exemplo, antígenos com especificidades totalmente diferentes para ambos os loci de um gene HLA.

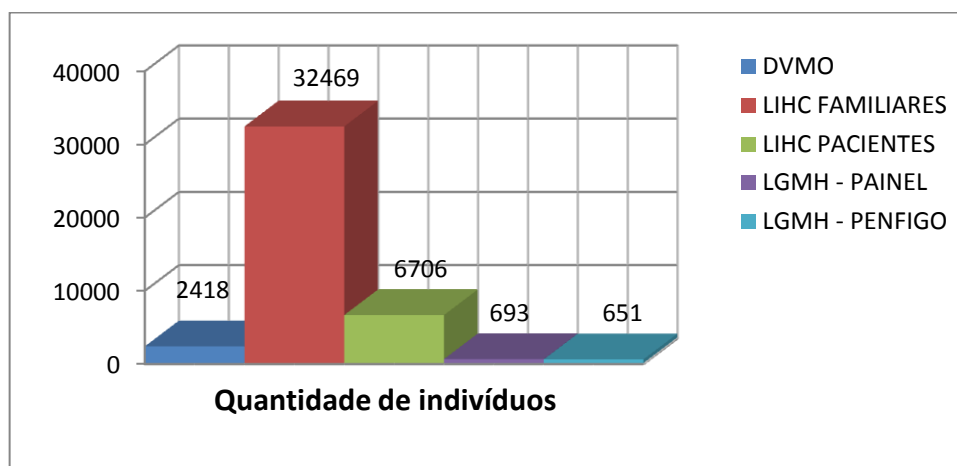


FIGURA 2 - Quantidade de indivíduos distribuído por origem dos dados

3.2 POPULAÇÃO DE ESTUDO

A população de estudo é composta por um conjunto de banco de dados distintos, sendo a origem destes dados o Hospital das Clínicas, Doadores Voluntários de Medula Óssea e o Laboratório de Genética Molecular Humana da UFPR. O primeiro conjunto a ser estudado trata-se do banco de dados do LIHC, ele é composto por pacientes do hospital e também por possíveis doadores familiares destes pacientes. A divisão desse conjunto de dados em grupos étnicos está

distribuída da seguinte forma: O conjunto de pacientes do HC é composto por 12 etnias distintas sendo 6706 indivíduos agrupados conforme

QUADRO 1:

QUADRO 1. Quantidade de indivíduos agrupados por etnia e origem

ETNIA	QUANTIDADE	
	PACIENTES	FAMILIARES
Branco	2198	11724
Asiático	22	95
Asiático/branco	1	7
Asiático/mulato	1	0
Cafuzo	10	1
Indiano	1	24
Índio	2	16
Mameluco	8	10
Mulato	494	2239
Mulato/índio	1	0
Não informado	3888	17783
Negro	57	465
Pardo	23	105

A proporção étnica entre esses dois conjuntos em alguns casos tem certa semelhança, futuras análises serão realizadas no decorrer deste trabalho.

O banco de dados de doadores voluntários de medula óssea é composto por 5 etnias distintas e um grupo que não informou a etnia durante a pesquisa. A divisão deste grupo encontra-se da seguinte maneira: asiático (n=33); branco (n=1851); índio (n=6); não informado (n=43); negro (n=111) e pardo n=(374), totalizando 2418 seres humanos. Apesar da quantidade inferior de indivíduos, este banco apresenta uma característica importante que é o campo CEP (Código de Endereçamento Postal), através dele é possível visualizar como está distribuída a população agrupando-os por tipagem HLA e grupo étnico. Este banco de dados é o mesmo que é encaminhado para o REDOME (Registro Nacional de Doadores de Medula Óssea). Os indivíduos deste banco de dados foram tipados no Hospital das Clínicas (HC).

Já os bancos cedidos pelo LGMH são divididos em dois grupos, os de pacientes com Pênfigo e um segundo grupo de Controle de indivíduos da região do Paraná. As amostras estão dispostas da seguinte forma, para o primeiro grupo asiático (n=1), branco (n=328), Branco brasileiro (n=17), branco/índio (n=3), índio (n=65), Mulato claro (Mais europeus que negros) (n=102), Mulato escuro (n=17), Mulato médio (n=63), mulato/índio (n=26), não informado (n=15), negro (n=13) e negro/índio (n=1) somando um total de 651. Para este conjunto de dados também foi possível obter a geolocalização com base no campo nascimento, o estudo visual relacionado a etnia e a tipagem HLA pode ser realizado. Já o segundo grupo é composto por 693 pessoas agrupadas conforme **FIGURA 3**. Gráfico comparativo etnias LGMH banco controle

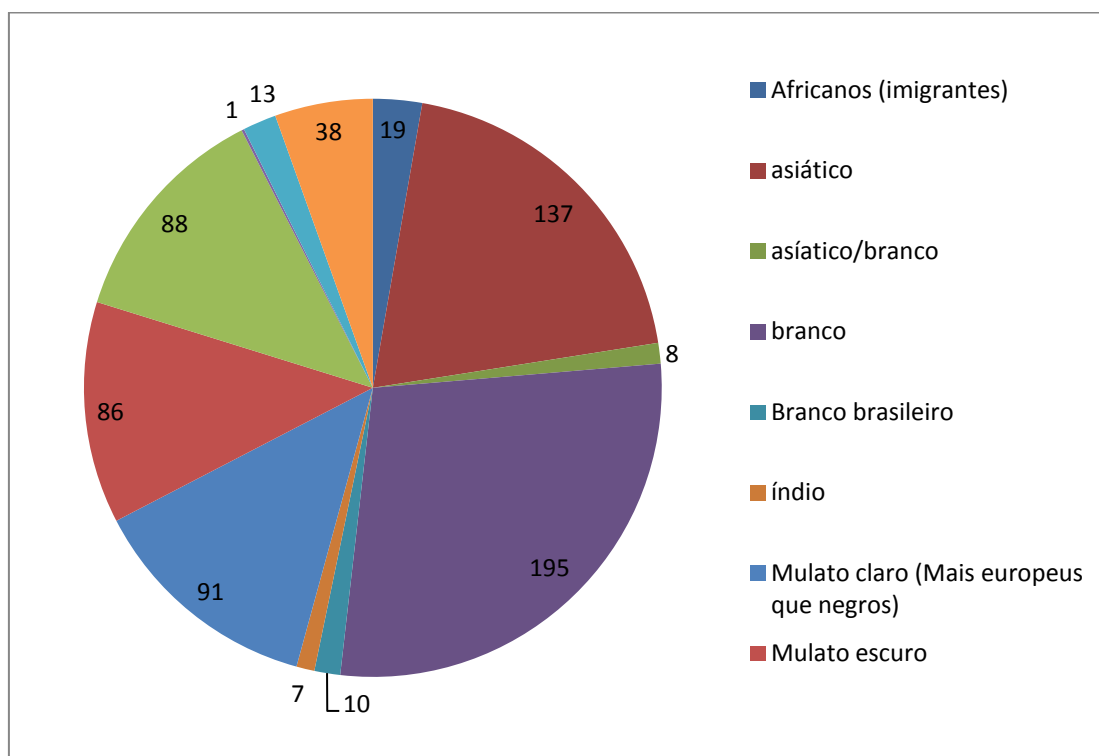


FIGURA 3. Gráfico comparativo etnias LGMH banco controle

O total de amostras representam 42937, sendo do sexo masculino (n=20821), ou aproximadamente 48,49% e do sexo feminino (n=21512) ou cerca de 50,10%, sendo que somente 1,40% ou (n=604) não apresentam sexualidade informada, isso demonstra uma homogeneidade dos dados, além do grande número de indivíduos o que permite estimativas e inferências estatísticas mais acuradas quanto em relação a população brasileira.

3.2.1 Propriedades das informações e confidencialidade

Este projeto de pesquisa foi encaminhado ao Comitê de Ética em Pesquisa em Seres Humanos do HC em 14 de julho de 2010 e teve sua aprovação e reconhecimento de que atende a aspectos das Resoluções CNS196/96 e demais, sobre Diretrizes e Normas Regulamentadoras de Pesquisa Envolvendo Seres Humanos do Ministério da Saúde. Está registrado sob CAAE: 0201.0.208.000-10 e Registro CEP: 2285.179/2010-07.

As informações foram utilizadas unicamente com finalidade acadêmica. Foram mantidos sigilo e cuidados exigidos com os dados, foi também omitida qualquer informação que permita identificação do paciente, doador, ou grupo estudado. Assim para realização dessa pesquisa não foi utilizado qualquer identificador de unicidade tais como Nome, Filiação, CPF, RG, Telefone e afins. Os componentes dos bancos foram identificados como indivíduos únicos por meio de código sequencial iniciando em 1 ou código de origem dos laboratórios envolvidos.

Os documentos encaminhados ao comitê podem ser encontrados na sessão de apêndice deste documento. Cabe ainda ressaltar o fato de que foram utilizadas somente duas únicas máquinas pessoais com o objetivo de preservar a segurança dos estados. Ambas eram protegidas com senha, tanto para entrada no Sistema Operacional como para acesso ao banco de dados. Mesmo que não fosse possível identificação do indivíduo do banco de dados, esta forma de trabalho foi adotada como pré-requisito para realização da pesquisa.

3.3 MATERIAIS UTILIZADOS

A classificação dos materiais utilizados está realizada em "*hardware*" e "*software*".

3.3.1 Do Hardware

Notebook

MARCA: AMAZON PC HEL-83

PROCESSADOR: INTEL CORE 2 DUO 2.0

MEMÓRIA RAM: 2GB DDR2 667MHZ

DISCO RÍGIDO: 120 GB SATA

Desktop

PLACA MÃE: ASUS

PROCESSADOR: AMD X2 5200

MEMÓRIA RAM: 8GB DDR2 800MHZ

DISCO RÍGIDO: 750 GB SATA

3.3.2 Do Software

Sistema Operacional Windows 7 Profissional

Microsoft Office 2007 Enterprise

ARLEQUIN 3.5.1.2 (Estatístico)

Pypop (Estatístico)

PostgreSQL (Banco de dados)

PHP 5 (Linguagem de Programação)

WAMP 2.Oi (Servidor WEB)

API Google Maps (Plotagem de dados)

WebService Correios (Tradução CEP)

3.3.2.1 Sistema operacional Windows 7 Profissional

Trata-se de um sistema operacional desenvolvido pela Microsoft. É responsável pela integração entre os "*softwares*" e os "*hardwares*" do computador. Este sistema foi utilizado em virtude da compatibilidade com todos os programas utilizados.

3.3.2.2 Microsoft Office 2007 Enterprise

É um pacote de aplicativos para escritório. Engloba editor de planilha eletrônica, editor de texto, banco de dados e criador de apresentações.

3.3.2.2.1 Access 2007

Este programa é utilizado para criação e controle de banco de dados. Foi desenvolvido pela Microsoft e apresenta-se como um sistema de fácil utilização para gerenciamento dos dados, criação de relatórios e armazenagem de informação ¹. Foi utilizado em um primeiro momento para acesso aos dados do LIHC e do LGMH.

3.3.2.2.2 Word 2007

É um editor de texto. Permite desenvolvimento e formatação de documentos profissionais². Foi utilizado para escrita da documentação deste trabalho.

3.3.2.2.3 Excel 2007

É uma ferramenta de análise sofisticada ³. Inclui uma variedade de maneiras de se gerenciar e analisar as informações. Foi utilizado no trabalho em um primeiro momento para agregação das informações. Após agregação dos dados e filtragem preliminar estes foram enviados para o banco de dados PostgreSQL. Esta ferramenta também foi utilizada para geração de gráficos e análises estatísticas através do suplemento extra do Office.

3.3.2.3 ARLEQUIN 3.5.1.2

¹ Acess - <http://office.microsoft.com/pt-br/access/recursos-e-beneficios-do-access-2010-HA101809011.aspx> . Acesso em 01/07/2011

² Word -
(http://www1.buyoffice.microsoft.com/bra/product.aspx?family=o14_word&country_id=BR&WT.mc_id=ODC_ptBR_Word_Buy) . Acesso em 01/07/2011

³ Excel -
(http://www1.buyoffice.microsoft.com/bra/product.aspx?family=o14_excel&country_id=BR&WT.mc_id=ODC_ptBR_Excel_Buy) . Acesso em 01/07/2011

Trata-se de um programa para análise genética de populações ⁴. Tem sua distribuição e uso gratuitos. Permite análise de equilíbrio de HardyWeinberg, cálculo de frequências alélicas e haplotípicas, entre outros.

3.3.2.4 PYPOP 0.7.0

Foi utilizado para análise das frequências haplotípicas. É uma ferramenta de uso e distribuição gratuitas, é escrita em Python e pode ser baixada em (<http://www.pypop.org/>). Permite análises semelhantes às do software Arlequin. O programa foi executado através de batch, em específico o arquivo: "pypop-batch.bat". Este arquivo foi executado em conjunto com arquivo de parâmetro configurado com as análises necessárias em cada momento.

Abaixo encontra-se um exemplo das análises realizadas. O parâmetro "-o" indica que será utilizado um arquivo de configuração diferente do valor padrão. Após a definição do arquivo de configuração de extensão ".ini", as populações a serem analisadas devem ser selecionadas. No exemplo abaixo, as populações dados_2A_1.pop e dados_2A_14.pop serão processadas.

```
C:\Users\LUIZ\Desktop\PyPopWin32-0.7.0>pypop-batch.bat -o config.ini  
dados_2A_1.pop dados_2A_14.pop
```

3.3.2.5 POSTGRESQL 9

A escolha pelo banco de dados PostgreSQL⁵ foi em virtude de ser um software livre, possuir um ótimo desempenho, ser multiplataforma funcionando em ambiente UNIX ou Windows, além de ser altamente escalável e possuir integridade referencial e suporte a transações (STONEBRAKER, 1986; STONEBRAKER, 1988;

⁴ Arlequin (<http://anthro.unige.ch/software/arlequin/>)

⁵ Postgresql (<http://www.postgresql.org>)

STONEBRAKER, 1990; STONEBRAKER, 1991; STONEBRAKER, 1975 e STONEBRAKER, 1992).

O PostgreSQL é um software banco de dados o qual armazena e organiza o conjunto de dados. A manipulação deste banco é dada através da linguagem SQL (*Structured Query Language*) ou Linguagem de Consulta Estruturada. Trata-se de uma linguagem pela qual é permitido inserir, alterar, deletar ou filtrar os dados de acordo com os padrões estabelecidos no código.

Basicamente foi a junção de três bancos de dados Access e uma planilha EXCEL, convertidos e ajustados para armazenagem no banco PostgreSQL. A **FIGURA 4** ilustra uma simplificação do processo ocorrido. Uma cuidadosa verificação manual foi realizada no banco, onde alguns dados sofreram alteração e alguns registros foram excluídos em virtude de divergência ou redundância.

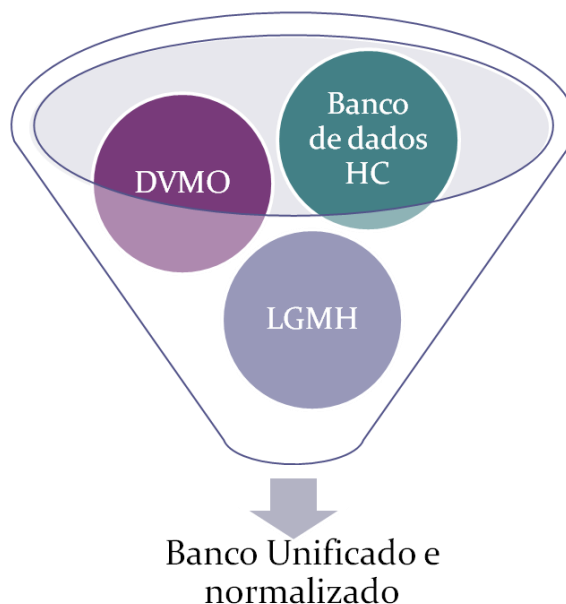


FIGURA 4 - Simplificação da unificação

Existem 4 operações básicas em SQL, elas são: Inserir, Atualizar, Deletar e Visualizar. A operação de inserir dados é realizada através do comando INSERT. A operação atualizar através do comando UPDATE, a de deletar através do comando DELETE e a de visualização dos dados através do comando SELECT.

A estrutura de um comando difere para cada tipo de operação desejada. No caso da operação de visualização dos dados a estrutura é a descrita abaixo:

```
SELECT * FROM tabela
```

O comando SELECT nos informa que iremos apenas visualizar os dados. Logo após o comando SELECT é opcional ao usuário visualizar todos os campos da tabela, no caso, para essa situação o símbolo "*", asterisco, é utilizado, ou ainda podendo expressar a visualização somente de determinados campos bastando mencioná-los logo após o comando SELECT. O código ficaria semelhante ao demonstrado abaixo:

```
SELECT campo1, campo2 FROM tabela
```

No caso demonstrado acima somente os campo1 e campo2 seriam retornados como resultado. O resultado seria algo semelhante ao demonstrado abaixo:

Campo 1	Campo 2
Dado A	Dado B
Dado C	Dado D
Dado E	Dado F

Outra característica dos comandos em SQL é que geralmente além de selecionar os campos desejados utilização a cláusula WHERE que nos permite visualizar somente dados com as características de filtro desejadas. Por exemplo, para visualizar somente dados cujo nome fosse "Maria" bastaria adicionar ao final do comando a cláusula "WHERE campo1="MARIA"".

Outro importante comando do SQL trata-se do UPDATE ele é realizado da seguinte forma:

```
UPDATE tabela SET campo1="X" campo2="Y" WHERE campo3="A"
```

No comando acima após o código UPDATE citamos a tabela que desejamos atualizar e em seguida o comando SET para dizer quais campo serão atualizados e quais informações eles receberão. No caso em questão o campo1 receberá "X" o campo2 receberá "Y" quando o campo3 for igual a "A". É extremamente necessário adicionarmos uma condição através do comando WHERE. Caso isso não ocorra todos os registros da tabela sofrerão a mesma atualização. Mais informações sobre as demais operações SQL podem ser encontradas no artigo (TIAGO, 2010)

Alguns dos comandos SQL realizados serão comentados, como por exemplo, a criação de tipagem com quantidade par de caracteres. Uma tipagem em baixa resolução para HLA-C como "3" , após o comando SQL é transformada em "03", para atender parcialmente a na nova nomenclatura adotada mundialmente e a criação de uma forma padrão para representação destes dados de forma única.

```
SELECT char_length(baixa_2c) as "Tamanho" , baixa_2c, idelemento FROM elemento
WHERE char_length(baixa_2c)=1
```

No exemplo acima foram selecionados somente registros que possuísem um caractere e a estes registros foi adicionado "0" manualmente. Porém aos que eram compostos por " * ", antes da tipagem foram desconsiderados em virtude de serem dados que apresentam dúvida em relação a sua tipagem. Tal processo foi realizado manualmente para que não ocorresse acidentalmente preenchimento incorreto dos dados e para avaliação de possíveis caracteres estranhos. Logo após o comando de atualização dos dados, indicado abaixo, era executado:

```
UPDATE ELEMENTO SET baixa_2c='0' || baixa_2c WHERE char_length(baixa_2c)=1 AND
baixa_2c!='0'
```

Neste comando os dados com tamanho de 1 e que são diferentes de "0", recebem um "0" a frente., frequentemente a inserção de filtros com cláusulas WHERE foi necessária para não ocorrer alterações indesejadas.

Outro comando realizado foi a criação de um padrão para campos não tipados, todos eles receberam o dígito "0". Antes dessa normalização os campos apresentavam-se com "0", "?", "NT", espaço vazio ou "*" os comandos executados para esta normalização foram:

```
UPDATE ELEMENTO SET baixa_2c='0' WHERE baixa_2c='*' OR baixa_2c='?' OR
baixa_2c='NT' OR baixa_2c=' '
```

Outro problema encontrado foi relativo às tipificações de alta e baixa resolução para o mesmo locus. Por exemplo, uma tipagem em alta resolução para o HLA-A sendo "0301" e "1511" e uma tipagem em baixa resolução sendo "04" e "03"

divergem entre si. Neste caso tais registros foram descartados. O comando abaixo foi executado para encontrar falhas desse tipo:

```
SELECT idelemento, alta_1a, alta_2a, baixa_1a, baixa_2a FROM ELEMENTO WHERE
(alta_1a!='0' and baixa_1a='0') OR (alta_2a!='0' and baixa_2a='0') OR (substring(alta_1a
from 1 for 2)!=baixa_1a AND alta_1a!='0') OR (substring(alta_2a from 1 for 2)!=baixa_2a
AND alta_2a!='0')
```

O quadro abaixo ilustra de forma resumida algumas das operações citadas acima:

QUADRO 2. Quadro comparativo de resultados obtidos no banco

BD ORIGINAL	BD UNIFICADO
?	0
NT	0
*	0
Espaço vazio	0
999	9999
9	99

O campo 999 ilustra uma tipagem de 3 caracteres que foi transformada em uma de 4, o mesmo raciocínio para o campo de 1 caractere diferente de 0 que passou a ser 2 caracteres. A criação desses padrões para representação dos dados, permite mais fácil comparação entre os diversos dados bem como uma estrutura mais confiável e de fácil operação e visualização.

3.3.2.6 PHP 5.2

O PHP ou *Hypertext Preprocessor* é uma linguagem de programação dinâmica e interpretada voltada para o desenvolvimento de páginas WEB. As possibilidades de aplicação do PHP são inúmeras e a opção por esta linguagem deu-se em virtude de não possuir restrição quanto a sistema operacional, pois funciona em ambiente UNIX e WINDOWS, tanto para cliente como servidor. O PHP, com o passar dos tempos, vem se mostrando uma ferramenta muito poderosa. Principalmente a partir de sua versão 5, mostra-se que é uma ferramenta de fácil uso, por exemplo, com o incremento da possibilidade de utilização de Orientação a

Objetos (OO) e Objeto de dados PHP (PDO), o que por fim permite adicionar uma camada de abstração que por sua vez facilita a criação de aplicações de negócio (PERRIN, 2006). O programa utilizado, que é responsável por servir as páginas WEB, é o APACHE⁶. Trata-se de um servidor WEB livre, o mais conhecido desde abril de 1996. Tem suporte HTTPS, que fornece certificados de conexão e sessão, realiza também criptografia dos dados, assim ofertando segurança para sistemas desenvolvidos e gerenciados pelo mesmo.

Os arquivos de conexão, entre o banco de dados e as páginas PHP, estão separados em arquivos distintos das páginas de apresentação de conteúdo. Existe o arquivo "inc.conectapdo.php" e o arquivo "inc.banco.php" que são responsáveis pela conexão. O primeiro corresponde à classe de conexão em si, e o segundo possui os dados de acesso, como endereço do servidor, banco de dados, usuário e senha. O usuário que acessa a página PHP via web browser não tem acesso a esses dados, o que preserva e garante, em partes, a segurança do banco de dados. A conexão com o bando de dados foi realizada a partir da interface de conexão PHP Data Objects (PDO), trata-se de uma camada de abstração dos dados que permite a conexão a diversos bancos⁷. Tal facilidade permite realizar uma conexão com diversos bancos distintos, desde que utilizando-se do driver, (mecanismo de conexão), apropriado, sem ter que se preocupar com sintaxe específica, além de se mostrar um mecanismo mais seguro de conexão não permitindo SQL Injections, dessa forma garantindo a confiabilidade e integridade dos dados.

3.3.2.7 WAMP 2.0i

É um pacote de software para Windows que engloba os programas Apache, MySQL, PHP, PhpMyAdmin e SqlBuddy. Neste trabalho foi utilizado o servidor Apache para hospedagem das páginas PHP e o PHP em si como interpretador de código. A versão do Wamp utilizado foi a 2.0, que engloba Apache 2.2.11 e PHP 5.2.9. Em um primeiro momento foi instala a versão PHP 5.2.5 para compatibilidade de conexão entre PHP e PostgreSQL, porém esta versão não é mais necessária. O

⁶ Apache (<http://www.apache.org>) Acesso em: 01/07/2011

⁷ Php: Hypertext preprocessor (<http://www.php.net>) Acesso em: 01/07/2011

código PHP atual utilizado nas páginas do sistema é totalmente compatível com as novas versões.

3.3.2.8 API GOOGLE V 3.0 e WebService dos Correios

É uma API escrita em linguagem JavaScript e distribuída pela GOOGLE. Esta linguagem de programação foi criada com o intuito de tornar dinâmico o conteúdo das páginas WEB e vem ganhando cada vez mais funcionalidades e poder de processamento (RAMOS *et al*, 2009). É processada do lado do cliente e interpretada pelo navegador, tal estrutura permite, teoricamente, que todas as páginas possam ser processadas independente de sistema operacional. A escolha da API do Google deu-se em virtude da constante atualização da mesma e a independência de sistema operacional ou instalador. Todo o código para plotagem dos dados é rodado no lado do cliente, acessando somente aos mapas do Google Maps. A documentação completa da linguagem pode ser encontrada no endereço: <http://code.google.com/intl/en-US/apis/maps/documentation/javascript/>. Uma das maiores vantagens dessa linguagem é, além da extensa documentação distribuída pela Google tanto nos idiomas inglês e português, é também o fato das comunidades criadas para discussão do assunto. Navegando pela internet é possível encontrar diversas aplicações para a mesma com diversos exemplos e códigos fontes das mais diversas naturezas.

Um dos problemas desta linguagem é a carência de padrão adotada pelos navegadores, sendo que alguns comandos específicos possam funcionar somente em algumas versões específicas de alguns navegadores. O esforço do Google é que a API funcione na grande maioria dos navegadores, um dos outros fatores para a escolha desta API.

A API do Google também foi utilizada como motor de conversão de endereço e nome de cidade em coordenadas de latitude e longitude, sem estes dados geográficos seria impossível visualizar os dados no mapa. Mais a frente na documentação será possível encontrar trechos dos scripts utilizados para tal processo.

A versão utilizada em questão foi a Version 3, compatível inclusive com sistemas operacionais baseados no Android e desenvolvida com o intuito de rodar

rapidamente em dispositivos móveis. Os serviços e manipulação dos dados na API do Google ocorrem naturalmente e foi, de certa forma, bem simples sua implementação, o que de fato em outras linguagens tomaria muito tempo na criação de módulos e camadas. Os principais elementos utilizados foram controle de marcadores, conversão de geolocalização e controles de mapa.

Já o Webservice dos Correios foi utilizado para conversão de dados de CEP em um endereço físico. Um script foi criado para obtenção destes dados, CEPs que eram inválidos foram manualmente processados pela API do GOOGLE. Alguns dos dados gerados pelo GOOGLE também eram inválidos, grande parte destes dados ocorreu em vista de erro de digitação do campo nascimento de alguns indivíduos. Estes indivíduos foram processados novamente, porém dessa vez uma pesquisa foi realizada com possíveis resultados e com o estudo do possível erro ocorrido no campo.

3.4 ESTATÍSTICA

Utilizando-se a técnica de máxima verossimilhança e de contagem de genes, foi obtida a quantidade, com a frequência e a distribuição dos alelos. O modelo de máximo verossimilhança é um procedimento que permite a vetores de parâmetros a realização de testes de hipóteses relativos a restrições lineares e não lineares, bem como a estimação dos parâmetros dos modelos econométricos. Um fato que deve ser levado em conta é o equilíbrio de Hardy-Weinberg, um parâmetro utilizado para provar que, mantidas as mesmas condições, ao passar das gerações, a frequência dos genótipos de um certo grupo se mantém. Ou seja, por intuição pensa-se que um gene raro ir-se-á a medida do tempo tornar-se mais raro e um gene mais comum ir-se-á tornar-se mais frequente, porém o equilíbrio de Hardy-Weinberg demonstra matematicamente que isso não ocorre. Ou seja, esse equilíbrio demonstra que mantidas as mesmas condições, ao passar das gerações, a frequência de um gene se mantém. Para o modelo algumas suposições de simplificação são necessárias, tais como, uma população muito grande, acasalamentos se dão ao acaso, todos os membros da população são igualmente férteis e não há efeito de seleção natural. Mais testes estatísticos e que serão realizados podem ser encontrados em (SANCHEZ-MAZAS *et al.*, 2002).

Todos os cálculos foram realizados utilizando o programa ARLEQUIN (EXCOFFIER *et al.*, 2005) e PyPop(LANCASTER *et al.*, 2007)(LANCASTER *et al.*, 2003) os demais programas como EXCEL, PHP, WAMP foram utilizados para geração de gráficos, formatação de dados, realização de cálculos e afins. As frequências alélicas e genotípicas foram obtidas através do algoritmo desenvolvido, as quais foram submetidas ao programa ARLEQUIN por meio do PyPop. A frequência de um alelo (FA) é obtida através da soma dos heterozigóticos por duas vezes a quantidade de homozigóticos, dividindo o resultado pelo dobro do total de indivíduos, visualizada em (1) (OLIVEIRA, 2007).

$$FA = (het + 2 hom)/2n \quad (1)$$

Onde:

het: Heterozigóticos

hom: Homozigóticos

n: Número de loci

A estatística da diversidade genética baseada no equilíbrio de Hardy-Weinberg foi calculada através da fórmula (2):

$$H = 1 - \sum_{i=1}^k P_i^2 \quad (2)$$

Onde *K* indica o número de alelos em um locus e *P_i* a frequência do alelo. As frequências observadas foram confrontadas contra o equilíbrio de Hardy-Weinberg através do teste de (GUO e THOMPSON, 1992). As frequências haplotípicas foram obtidas através do algoritmo ELB, bem como os parâmetros de desequilíbrio de ligação utilizando método do qui-quadrado com correção de Yates.

A análise da possibilidade de utilização do banco de familiares de doadores como banco controle será comprovada através da análise multivariada. Uma tabela de contingência de 2x2 será criada e o teste do qui-quadrado será realizado para comparação. A aplicação da correção de Yates será utilizada somente quando necessária. Esta análise é semelhante a realizada por (SMOUSE E WILLIAMS, 1982). O quadro abaixo ilustra as fórmulas para cálculo do qui-quadrado, tanto para a fórmula de Karl Pearson quanto para de Frank Yates.

QUADRO 3. Tabela Qui-Quadrado

Mark Pearson	$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(A_{ij} - E_{ij})^2}{E_{ij}}$			
Frank Yates	$\chi^2 = \frac{N(a * d - b * c - \frac{N}{2})^2}{(a + c) * (b + d) * (c + d) * (a + b)}$			
Tabela				
		Alelo i		
		Presente	Ausente	Total linha
Alelo j	Presente	a	b	a+b
	Ausente	c	d	c+d
	Total Coluna	a+c	b+d	N=a+b+c+d

3.5 A CRIAÇÃO DO BANCO UNIFICADO

Para o desenvolvimento desse trabalho foram tomados diversos passos. Foram realizadas diversas análises, testes, filtragens tanto manuais quanto automáticas, construção de scripts entre outros. A figura abaixo demonstra resumidamente os passos realizados durante o desenvolvimento do trabalho.

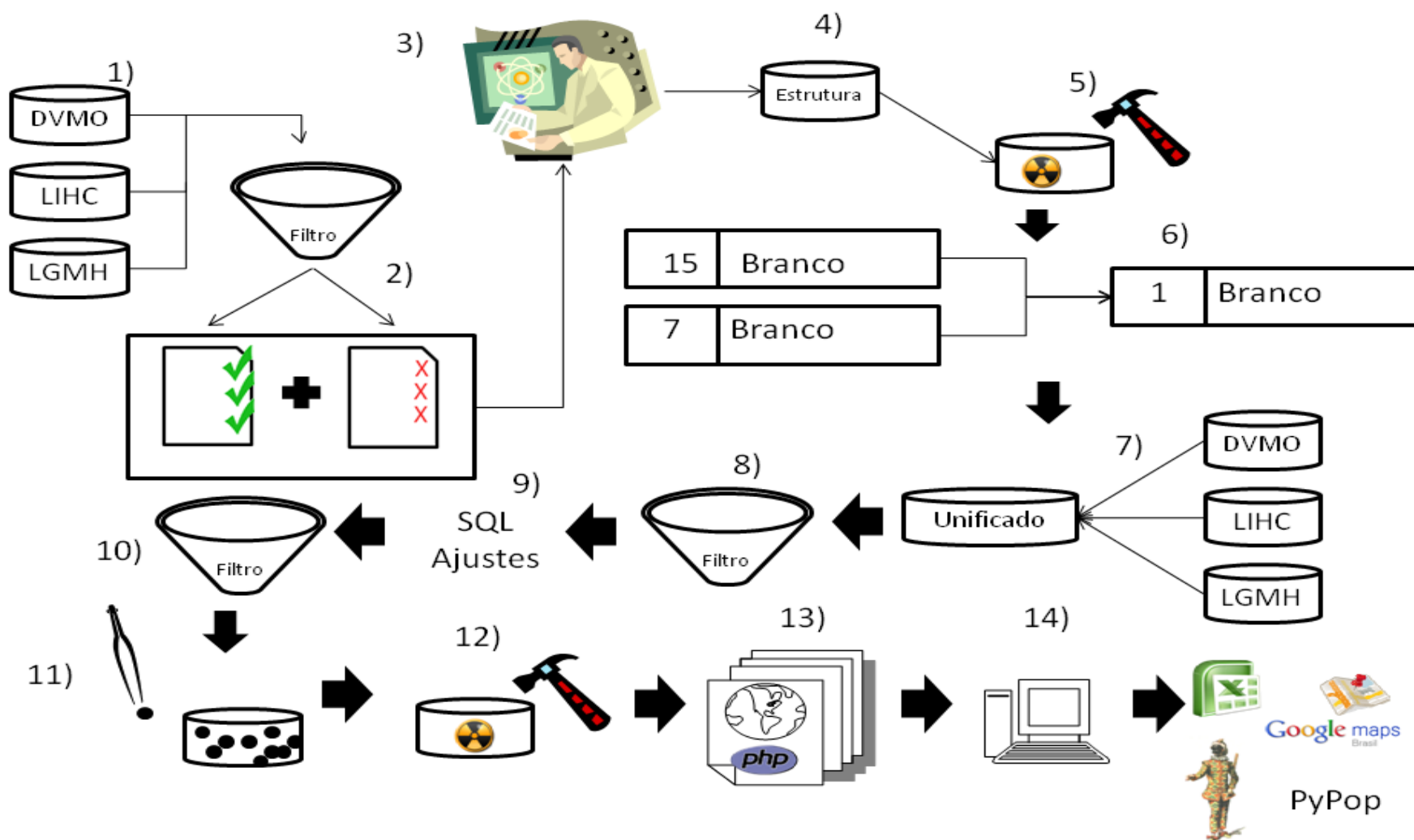


FIGURA 5 - Fluxo de trabalho

Inicialmente foi realizado um estudo dos dados que englobavam a pesquisa. Também foi realizada uma análise abordando os três bancos de dados. Estudou-se os campos dos bancos de dados que depois foram submetidos a um filtro onde separou-se os campos em comum entre os bancos e os divergentes, conforme passo 2. Novamente após uma nova análise dos dados obtidos, criou-se uma estrutura de banco de dados que abordaria todos os 3 bancos juntos. Esta estrutura foi testada para que se contemplasse todos os dados e que fosse possível realizar análises futuras. Procurou-se manter parte da estrutura antiga dos bancos e a criação do mesmo levou em conta a facilidade para extração de dados futuros. No passo 6, foram realizados agrupamentos e padronizações dos dados, pois para cada banco, por exemplo, a etnia branca possuía um código diferente. Outro exemplo foi a etnia asiática e amarela que foram agrupadas no mesmo grupo.

No passo 7 foram criadas estratégias de importação desses dados. Os três bancos estavam em fontes distintas, os bancos LGMH e LIHC estavam em formato de arquivo Access, enquanto o banco de DVMO estavam em planilha eletrônica do Excel. Os três bancos foram importados inicialmente pro formato de planilha e depois importados para o banco PostgreSQL através de scripts de automação.

Após criação do banco unificado os dados foram submetidos a filtros para separação de alguns ruídos existentes, tais como campos com caracteres inválidos entre outros. Comandos SQL de ajustes dos dados foram executados para criação de uma padronização na representação das informações. Alguns dos comandos foram descritos na sessão 3.3.2.5. Novamente os dados foram submetidos a filtros, onde foram analisados campos de observações dos bancos de dados. Tais campos envolviam anotações de usuários tais como "Mesmo cadastro de XXXXX", "Filho de XXXX", e assim por diante. Tais campos foram analisados e certas deleções foram realizadas para não haver duplicidade de informação. Ajustes também foram realizados com base nas informações obtidas desses campos. Outros ajustes manuais envolveram comparação entre indivíduos duplicados, dados divergentes tais como tipagem em alta e baixa resolução não correspondentes entre outros.

No passo 12 foram realizados mais testes no banco criado. Procurou-se atender todas as funcionalidades pretendidas. No passo 13 foram criados scripts para extração e representação dos dados. Estas páginas e scripts foram hospedadas por meio de um webserver onde em seguida os dados foram extraídos

e submetidos aos programas e api(Excel, Arlequin, Pypop e Googlemaps) para realização da pesquisa.

4 RESULTADOS

Os resultados abaixo estão divididos em acordo com os objetivos específicos, cada sessão será comentada independentemente.

4.1 FLUXO DE DADOS - O SISTEMA CRIADO E O BANCO DE DADOS

As páginas do sistema são hospedadas em um servidor. A localização do servidor do banco de dados e da aplicação não precisam necessariamente compartilhar o mesmo endereço físico. Ao administrador basta alterar o arquivo "inc.banco.php", do servidor de aplicação, com os dados do servidor de banco, inserindo no arquivo o endereço IP, o usuário e o nome do banco. As páginas são gerenciadas pelo servidor APACHE, este permite com que o cliente acesse o sistema e possa realizar as diversas funções criadas. O cliente acessa as páginas desenvolvidas em linguagem PHP por qualquer navegador, desde que este, no caso da API DO GOOGLE, tenha total compatibilidade com a versão 3.0 e habilitado o JavaScript. Ao usuário além da plotagem, geração de arquivos para o Arlequin, também permite a exportação dos dados para Excel, permitindo que este possa utilizar os dados em outros programas e realizar outras análises. A **FIGURA 6**, ilustra bem os processos de acesso do sistema.

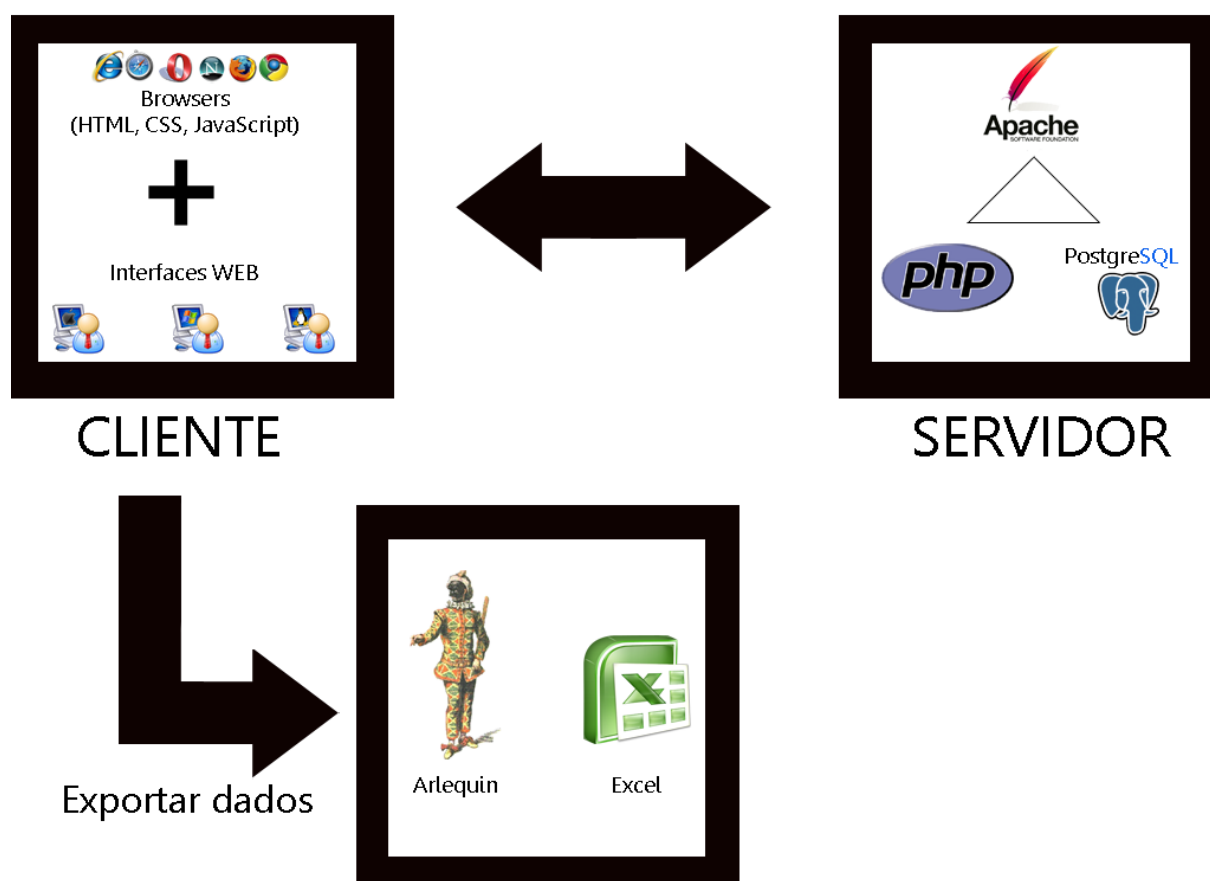


FIGURA 6 - Fluxo de dados

A **FIGURA 7** ilustra resumidamente os passos necessários para desenvolvimento deste trabalho e das páginas criadas. Inicialmente para responder ao problema de pesquisa os três bancos de dados (DVMO, LIHC e LGMH) foram submetidos a um filtro, onde através de comandos SQL automáticos e verificação manual foram eliminadas duplicidades, dados divergentes e criada uma padronização, isso para que futuras comparações pudessem ser realizadas sem maior dificuldade. Logo após um novo banco unificado foi criado com estes dados já filtrados. A partir deste banco unificado foram desenvolvidas páginas em linguagem PHP que são responsáveis por: Exportar dados para os softwares Arlequin e Excel, isso para cálculos de frequências alélicas e haplotípicas e também uma página responsável pela plotagem dos dados de tipagem no mapa, esta utilizando a API do GoogleMaps, que será comentada nas próximas sessões deste trabalho. Os dados então após submetidos a estas páginas geraram vários resultados que foram analisados nesse trabalho.

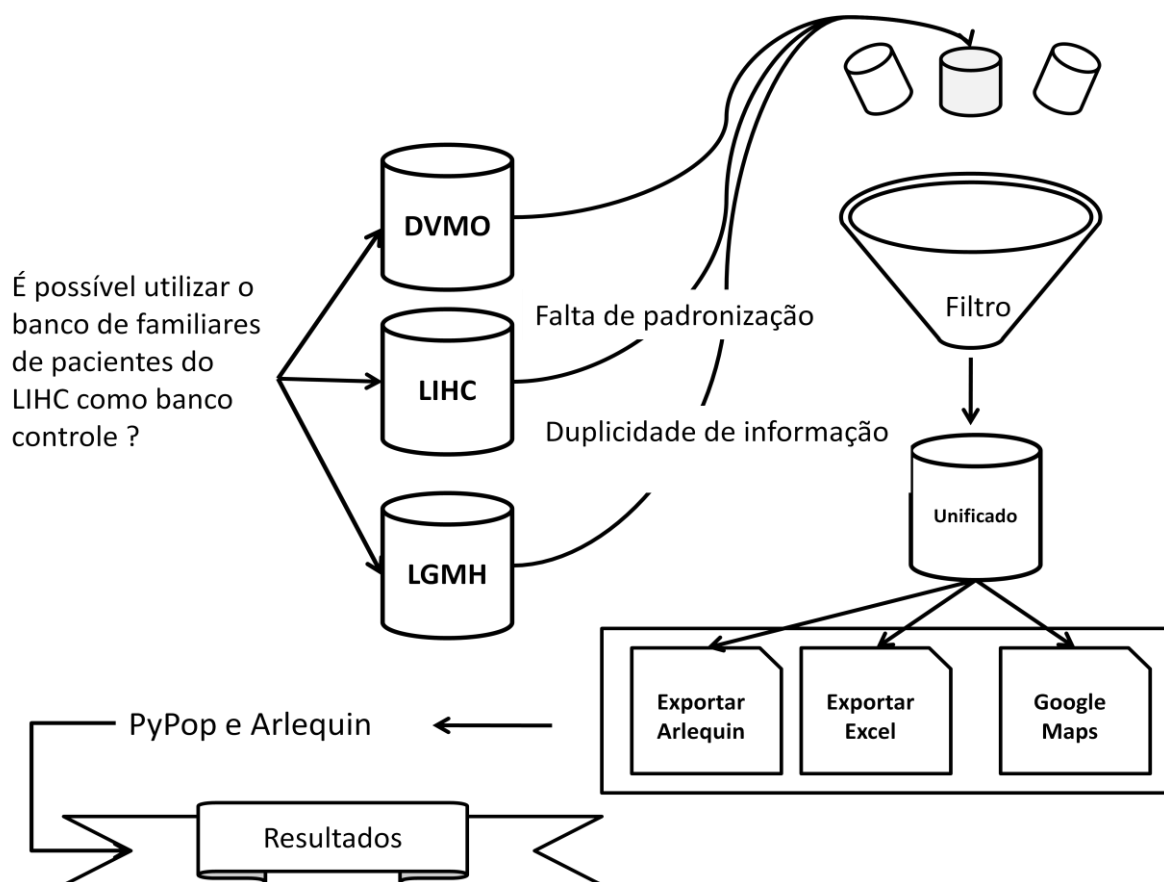


FIGURA 7 - Passos realizados no desenvolvimento

A **FIGURA 8** ilustra os casos de uso das páginas desenvolvidas. É possível para o usuário exportar os dados para o programa Excel, por meio de arquivo de planilha eletrônica ".xls" ou para o programa Arlequin por meio do arquivo ".arp". Ainda também é possível ao usuário plotar os dados de tipagem utilizando a API do GoogleMaps. Ambos os métodos são passíveis de aplicação de filtros de seleção de conteúdo desejado. A **FIGURA 9** ilustra melhor a etapa de aplicação de filtros. Através deles é possível selecionar a origem dos dados, as etnias ou ainda os genes HLA e locus desejados.

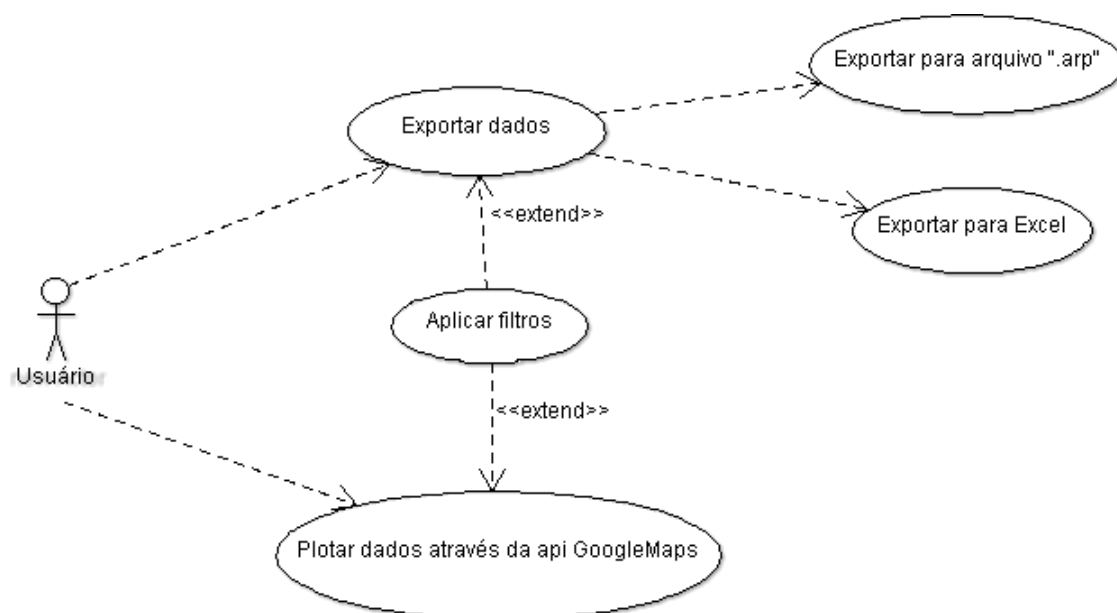


FIGURA 8 - Caso de uso páginas desenvolvidas

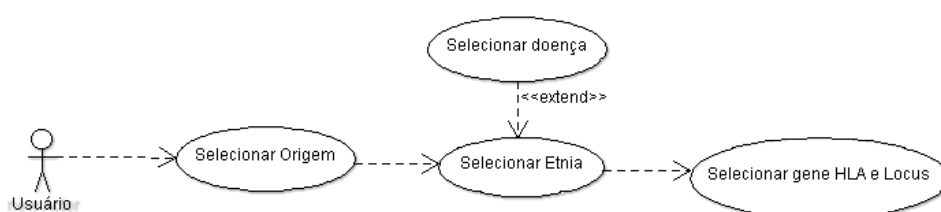


FIGURA 9 - Caso de uso exportar dados

Na **FIGURA 10** é possível ver a representação do banco de dados através do Diagrama Entidade Relacionamento. A principal tabela recebe o nome de elemento, ela contém basicamente todos os dados dos indivíduos. Esses dados englobam dados referentes a geo-posicionamento, tipagem HLA, grau de parentesco, origem do dado, doença e etnia. Os indivíduos que possuem grau de parentesco são relacionados através do Reg HC e Reg LHMH, cada um deles é um representante da família em questão. O campo código familiar representa o grau de parentesco com o indivíduo portador de alguma patologia. A tabela Etnia apresenta mais de 50 etnias. Algumas padronizações foram estabelecidas para que a comparação entre

os três bancos ocorresse de forma natural. Por exemplo, etnia asiática em um dos bancos foi substituída por amarela.

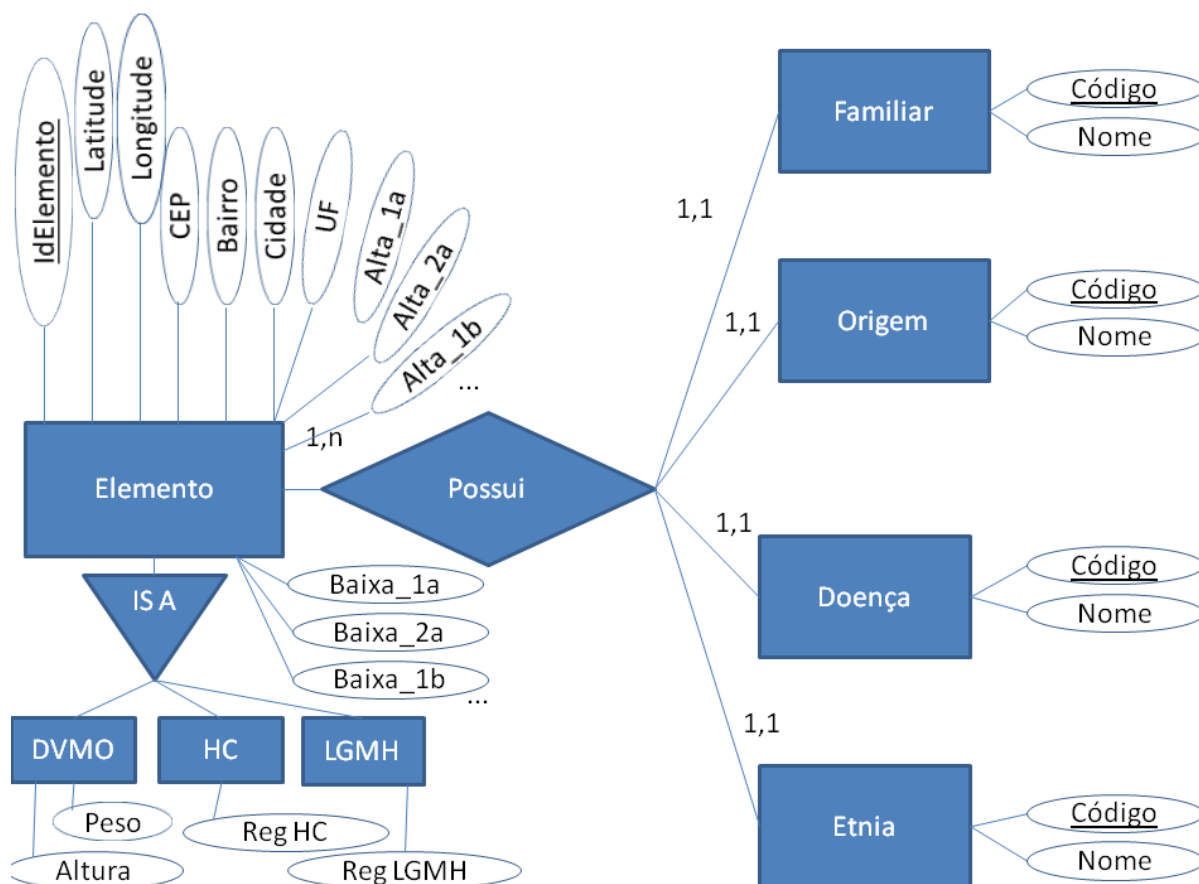


FIGURA 10 - Diagrama Entidade Relacionamento

4.1.1 Gerador de .Arp

Para criação dos dados estatísticos foi desenvolvido uma página WEB em PHP (Hypertext Preprocessor), a qual gera arquivos .ARP que são utilizados pelo programa ARLEQUIN. A seguir são demonstradas as etapas para obtenção dos arquivos organizados para estatísticas no programa.

A) Gerar arquivo .arp

List Box - Multiple Select

PAINEL Qtde: 693	<input type="button" value="Passo -> 2"/>
DVMO Qtde: 2418	
HC DOADORES Qtde: 32477	



B)

Gerar arquivo .arp

HC PACIENTES não informado Qtde: 3888	?	<input type="button" value="Passo -> 3"/>
HC PACIENTES branco Qtde: 2198	não informado	
HC PACIENTES mulato Qtde: 494	Linfoma	

C)

Gerar arquivo .arp

PAINEL branco Qtde: 195	<input type="button" value="Passo ->"/>
PAINEL asiático Qtde: 137	
PAINEL Mulato claro (Mais europeus que negros) Qtde: 91	



D)

Gerar arquivo .arp

Título:	Insira um título da busca	
Escolha a tipagem a ser verificada	alta_1a(2)	
Realizar para tipagem:	<input checked="" type="radio"/> Especifica 1	<input type="radio"/> Dupla
	<input type="radio"/> Especifica 2	
<input type="button" value="Gerar Arquivo"/>		

FIGURA 11 - Passos de criação de arquivo ".arp"

Na **FIGURA 11** sessão A, o usuário seleciona qual a origem dos dados, é permitido a este que selecione mais de uma origem. Os próximos passos são B ou C, na sessão C, o usuário seleciona as etnias que deverão estar presentes no estudo, caso seja selecionada na sessão A a origem PACIENTES, então a sessão B é carregada e um segundo listbox, campo que permite seleção múltipla, com as opções de seleção de tipos de doença é carregado.

Na **FIGURA 11** sessão D a frente do campo de escolha de tipagem, para as tipagens de alta resolução pode-se realizar análise envolvendo somente 2 ou 4 primeiros caracteres com objetivo de se eliminar dados estatisticamente irrelevantes. Por exemplo, ao selecionar a tipagem com o marcador (2), apenas os dois primeiros caracteres da tipagem selecionada serão levadas em consideração, o mesmo é válido para o número (4). As tipagens que não apresentam nenhum número a frente do nome consideram-se todos os caracteres do campo selecionado. O arquivo .arp pode ser gerado de 3 maneiras, sendo elas "Específica 1", "Específica 2" e "Dupla". No caso de seleção de "Específica 1" o arquivo gerado será o cálculo do grupo alélico selecionado, porém somente para o locus específico. A seleção do "Específica 2" retorna o arquivo avaliando os dois loci, tanto o 1 como o 2. Já a terceira opção, realiza um agrupamento analisando os indivíduos nos dois loci ao mesmo tempo. Nesse caso, a ordem dos valores não é importante pois o script organiza os dados e os agrupa, sendo assim uma tipagem HLA-A_1 com valor "01" e HLA-A_2 com valor "02", será agrupado com outro indivíduo HLA_A_1 com valor "02" e HLA_A_2 com valor "01".

O **QUADRO 4** é um descritivo dos campos do cabeçalho do arquivo gerado:

QUADRO 4. Descrição cabeçalho arquivo ".Arp"

CAMPO	DESCRIÇÃO
Title	Título da pesquisa
NbSamples	Número de amostras no arquivo
DataType	O tipo de dado que está no arquivo
LocusSeparator	O caractere separador de dados
MissingData	Dados omissos são preenchidos com o caractere selecionado
RecessiveData	Dados recessivos são preenchidos com o caractere selecionado

Nos campos referentes aos dados das amostras temos: SampleName, onde o usuário insere o nome da amostra, o campo SampleSize que apresenta o tamanho da amostragem e o campo SampleData, que apresenta os dados contabilizados pelo algoritmo. A **FIGURA 12**, abaixo, apresenta uma breve descrição.

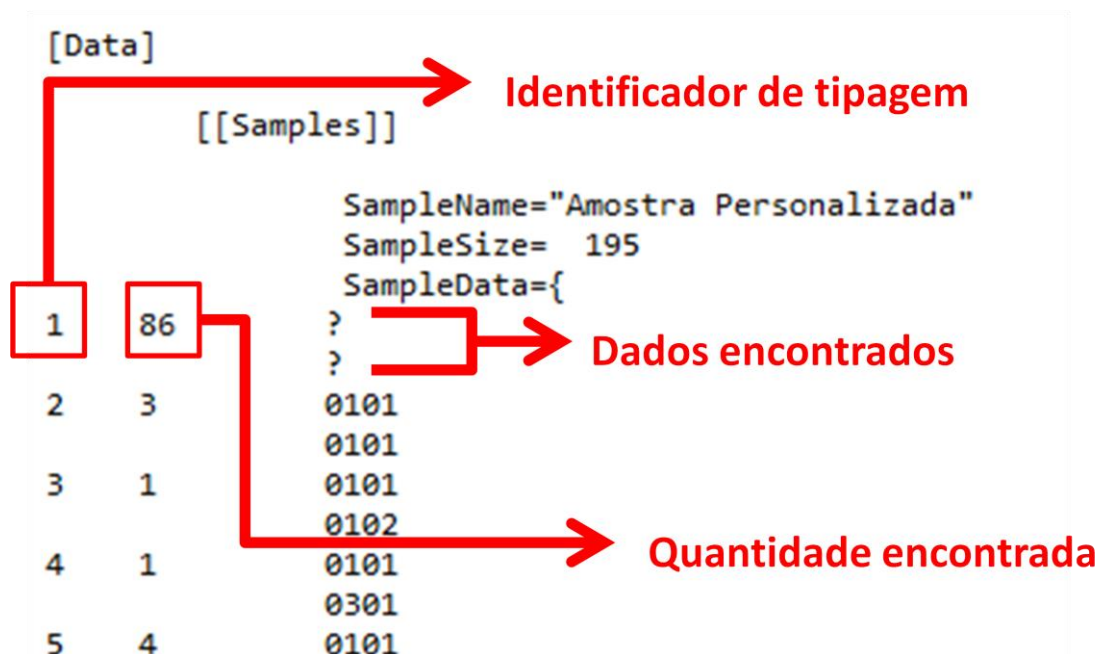


FIGURA 12 - Descrição arquivo ".arp"

O indicador de novo dado é a próxima linha de identificador.

A **FIGURA 13** e a **FIGURA 14** ilustram o arquivo ARP gerado a partir da seleção do usuário. A **FIGURA 14** é um arquivo que de análise envolvendo os dois loci, enquanto a **FIGURA 13** somente um deles. Por fim a **FIGURA 15** demonstra um arquivo de análise dos dois loci em conjunto.

[Profile]

```
Title="Insira um titulo da busca"
NbSamples=1
```

```
GenotypicData=1
GameticPhase=0
DataType=STANDARD
LocusSeparator= WHITESPACE
MissingData='?'
RecessiveData=X
```

```
#Try different epsilon value (in General Settings dialog box)
#and different values of No. of initial conditions
#(in the Diversity Indices dialog box)
#for the estimation of haplotype frequencies
#It should have an impact on the final likelihood of the sample

RecessiveAllele="null"
```

[Data]

[[Samples]]

```
SampleName="Amostra Personalizada"
SampleSize= 195
SampleData={
```

```
1 1 1201
2 1 1406
3 3 0407
4 2 1402
5 18 1501
6 2 1502
7 16 0101
8 1 0102
9 3 0301
10 2 1602
11 4 1301
12 3 1302
13 86 ?
14 4 0411
15 5 0801
16 1 1104
17 5 1601
18 2 0408
19 5 1503
20 2 0401
21 1 0403
22 5 1401
23 10 0701
24 1 0402
25 9 1101
26 1 0807
27 1 0804
28 1 1102
}
```

FIGURA 13 - Especifica 1 .arp

[Profile]

```
Title="Insira um titulo da busca"
NbSamples=1
```

```
GenotypicData=1
GameticPhase=0
DataType=STANDARD
LocusSeparator= WHITESPACE
MissingData='?'
RecessiveData=X
```

```
#Try different epsilon value (in General Settings dialog box)
#and different values of No. of initial conditions
#(in the Diversity Indices dialog box)
#for the estimation of haplotype frequencies
#It should have an impact on the final likelihood of the sample
```

```
RecessiveAllele="null"
```

[Data]

[[Samples]]

```
SampleName="Amostra Personalizada"
SampleSize= 195
SampleData={
01    1    (1401)
2    172    ?
3    20    0101
4    2    0102
5    18    0301
6    4    0401
7    1    0402
8    3    0403
9    1    0405
10   2    0407
```

FIGURA 14 - Específica 2 .arp

[Data]

[[Samples]]

```
SampleName="Amostra Personalizada"
SampleSize= 195
SampleData={
1    86    ?
2    3    ?
3    1    0101
4    1    0101
5    4    0101
0101
0101
0101
0102
0101
0301
0101
```

FIGURA 15 - Dupla .arp

Seguindo pela mesma idéia é possível também gerar arquivos com os haplótipos. A página responsável por gerar esses dados difere somente por seleção de 2 ou 3 genes distintos, conforme **FIGURA 16**. Por exemplo, é possível realizar o cálculo do haplótipo avaliando-se por simples ou duplo. No caso de escolha do haplótipo simples é realizado o agrupamento ordenando-se as tipagens e contando cada indivíduo uma única vez, a **FIGURA 24** do APÊNDICE 6, ilustra um resultado desse tipo. Para o haplótipo duplo são realizadas diversas combinações possíveis de fenótipo, desta forma temos um mesmo indivíduo sendo contado mais de uma vez como pode-se observar na **FIGURA 25** também do APÊNDICE 7.

Gerar arquivo .Arp

Título:	Insira um título da busca	
Escolha as tipagens a serem verificadas	<div> <div>alta_1a(2)</div> <div>alta_1a(4)</div> <div>alta_1a</div> <div>alta_1b(2)</div> <div>alta_1b(4)</div> <div>alta_1b</div> <div>alta_1c(2)</div> <div>alta_1c(4)</div> <div>alta_1c</div> <div>alta_1drb1(2)</div> </div>	
Tipo grupo: :	<input checked="" type="radio"/> Haplótipo Simples	<input type="radio"/> Haplótipo duplo
Gerar Arquivo		

FIGURA 16 - Gerar arquivo de haplótipo

4.1.2 Plotagem de dados

As aplicações de Sistemas de Informação Geográficas (SIG) vem se tornando componente importante na agregação, análise e desenvolvimento de dados geográficos. As aplicações práticas vão de estudos de estatística descritiva a evolução dos estudos de relação entre exposição a fatores ambientais e saúde da população (SEVER, 2004). O estudo da medicina geográfica é muito importante pois tem como objetivo prover entendimento dos problemas de saúde e melhorar a condição de saúde das pessoas com base nos fatores geográficos que influenciam

(BRINEY, 200-?). Existem várias aplicações de mapa tais como Google Flu Trends, WHO Map, Gallery from Word Health Organization, Index of Pain entre outros.

Uma página em Web utilizando a linguagem de programação PHP foi desenvolvida para plotagem dos dados no mapa geográfico. A plotagem ocorre por meio da API do GOOGLE MAPS. Para obtenção dos dados geográficos utilizou-se o CEP dos indivíduos do banco DVMO e para o banco PENFIGO e PAINEL utilizou-se o campo local de nascimento. Para conversão do CEP em um endereço válido um WEB SERVICE dos CORREIOS foi acionado, que em seguida o endereço foi submetido a API do GOOGLE para conversão em coordenada geográfica. O mesmo se deu para os bancos de PENFIGO E PAINEL do LGMH, o qual o campo nascimento foi convertido em coordenadas de latitude e longitude diretamente pela API do GOOGLE. A **FIGURA 17** demonstra uma exemplificação do processo.

O geo-processamento é importante principalmente para entender como os dados estão distribuídos e para gerenciar os recursos de, por exemplo, campanhas de saúde principalmente quando recursos para estes são escassos. Ferramentas SIG podem revelar dados que são difíceis de serem visualizados quando estão em forma de texto/tabela. A idéia de criação desse mecanismo era de através da visualização espacial da distribuição dos genes da população pudéssemos conseguir notar algum padrão. Padrão esse que poderia facilitar a busca de um doador ou simplesmente mostra-se como uma forma de compreendermos melhor a distribuição desses genes e a associação deles com os grupos étnicos. Ainda que trata-se de uma ferramenta simples os primeiros passos foram dados. Técnicas de agrupamento e relação com fatores ambientais podem ser empregadas no futuro, o principal que trata-se da tradução do endereço em coordenada geográfica já foi realizado.

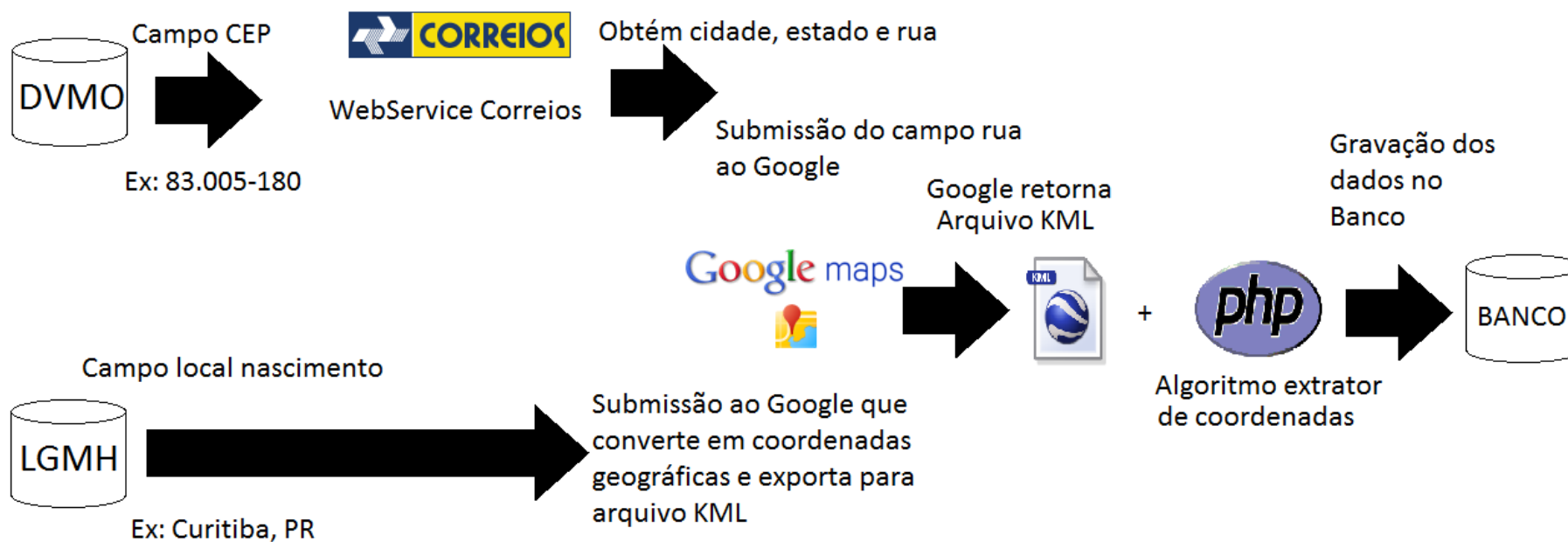


FIGURA 17 - Obtenção de coordenadas geográficas

Os passos para plotagem dos dados no mapa são praticamente os mesmos passos para criação do arquivo ARP já demonstrado anteriormente, porém no caso de plotagem é possível realizar a escolha de uma tipagem específica, por exemplo o valor "02" para o gene HLA-A. A **FIGURA 18** demonstra as possíveis configurações de plotagem dos dados.

Gerar Mapa

Escolha a tipagem a ser verificada	alta_1a(2) ▼
Realizar para tipagem: <input checked="" type="radio"/> Especifica Loci	<input type="radio"/> Dupla
Procurar: <input style="width: 100px;" type="text" value="0"/>	<input type="radio"/> Especifica
<input type="button" value="Ver no mapa"/>	

FIGURA 18 - Configuração plotagem

Na **FIGURA 19** é possível observar os dados plotados no mapa. Uma lista a direita do mapa é carregada com os indivíduos plotados. Ao clicar tanto nos marcadores quanto na lista a direita um balão informativo é aberto com os dados daquele indivíduo. É possível clicar no link VER PERFIL e observar mais dados sobre o objeto selecionado. As cores e letras dos marcadores são atribuídas através do agrupamento de ORIGEM, ETNIA e TIPAGEM. Os marcadores para distinção entre si são numerados de 1 a 260 recebendo caracteres de A a Z com cores diferentes. Os caracteres do alfabeto foram inseridos nos marcadores somente para distinção dos grupos, as letras do alfabeto multiplicadas pelas 10 cores existentes somam esse total de 260. Sendo assim, este sistema classifica corretamente 260 grupos distintos, caso o número de grupos existentes seja superior a 260, então o 261º grupo será agrupado com o grupo de numero 1 e assim por diante.



FIGURA 19 - Plotagem Simples

Na **FIGURA 20** é possível observar que abaixo do mapa é carregado diversos checkbox, estes checkbox são no número exato de grupos classificados de acordo com a seleção do usuário do sistema. Este controle permite ao usuário uma filtragem dos dados para melhor visualização e comparação bem como reconhecimento de padrões.

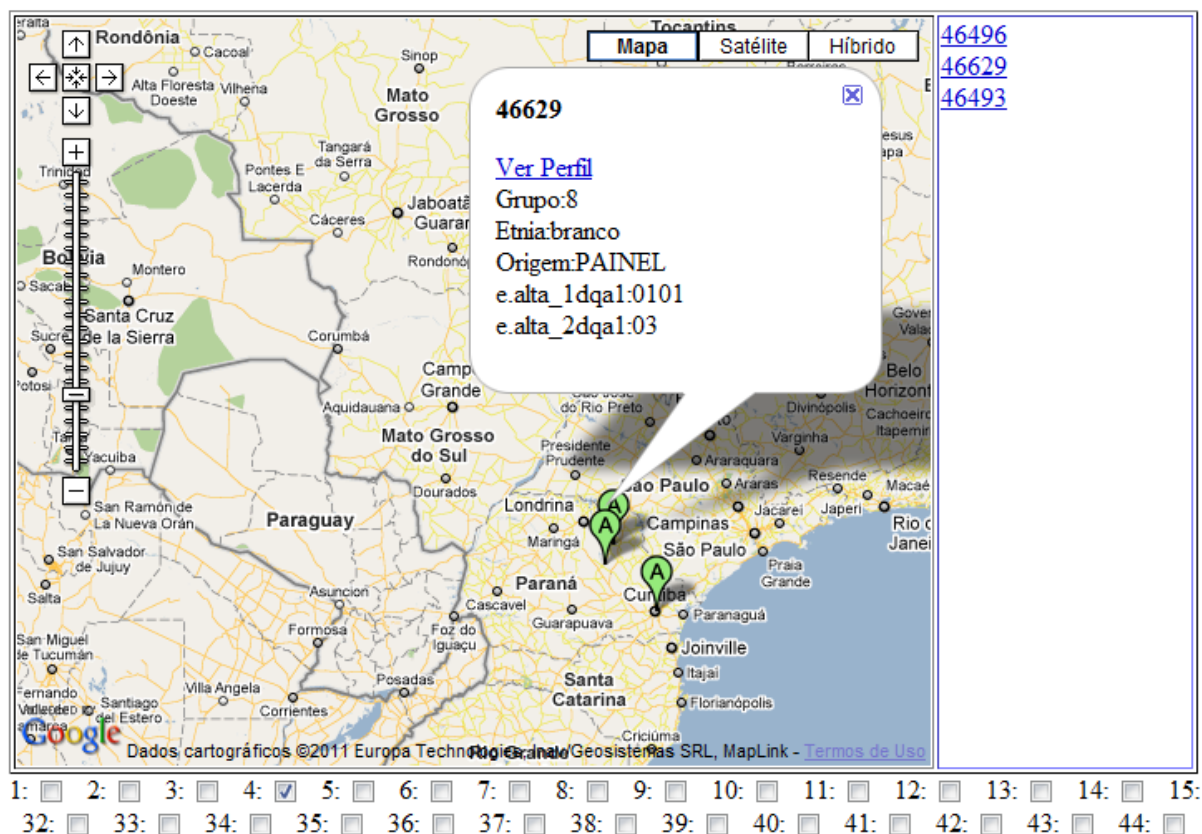


FIGURA 20 - Plotagem dupla loci

4.1.3 Exportação de dados para o Excel

Uma das páginas criadas tem a função de exportação dos dados para o formato de planilha eletrônica com extensão ".xls". O arquivo gerado por esta página pode ser aberto no Microsoft Office Excel, Open Office ou ainda Google Docs, sendo totalmente compatível com qualquer uma das versões destes programas. A **FIGURA 21** ilustra a tela de exportação destes dados. Se o usuário seleciona a opção "Códigos", este para os campos etnia, doença, origem e grau de parentesco irá somente apresentar os identificadores únicos do banco de dados. Por exemplo, para a etnia branca, o identificador do banco de dados é o número 1, sendo assim no, arquivo exportado, para o campo etnia seria visível somente este número. Na seleção da opção nomes, o arquivo exportado recebe a tradução do identificador único do banco no nome específico. Para o exemplo citado acima, ao invés de visualizar o identificador número 1, seria visualizado etnia "branco". Já para a

seleção da opção "ambos" seriam exportados ambos os campos, o da nomenclatura e o do identificador único. Este recurso foi criado em especial para que os usuários que não desejem trabalhar com a interface criada possam exportar esses dados para outros programas e realizar, assim, outros diversos estudos não contemplados neste trabalho. Esse procedimento foi realizado para trabalhar com o programa PYPOP.

Para uso no programa PYPOP alguns ajustes foram necessários por exemplo, deleção de todo o conteúdo que não fosse desejado ser analisado pelo programa. Ou seja, colunas que não fossem interessantes na análise foram deletadas, ainda no caso dos antígenos para os campos com tipagem vazia, ou seja caractere '0', foram substituídos por '****', conforme padronização do PYPOP.

Exportar Excel

Selecione uma ou mais origem
de dados

☐ Codigos
☐ Nomes
☒ Ambos

PAINEL Qtde: 693	▲
DVMO Qtde: 2418	■
HC DOADORES Qtde: 32477	▼

BD -> Excel

FIGURA 21 - Criação de arquivo de planilha eletrônica

A figura baixo ilustra um arquivo gerado com a opção "Ambos" selecionada. É possível visualizar claramente os campos etnia, origem e doença sendo traduzidos os identificadores nas nomenclaturas correspondentes.

ELEMENTO	ETNIA	ETNIA	DOENCA	DOENCA	ORIGEM
46818	asiático	6	Livre	54	PAINEL
46532	Mulato médio	18	Livre	54	PAINEL
46540	branco	1	Livre	54	PAINEL
46646	Mulato claro (Mais europeus que negros)	17	Livre	54	PAINEL
46538	branco	1	Livre	54	PAINEL
46641	branco	1	Livre	54	PAINEL
46479	branco	1	Livre	54	PAINEL

FIGURA 22 - Dados exportados

4.2 ANÁLISES ESTATÍSTICAS

Nas sessões que se seguem serão demonstrados os resultados obtidos através da análise dos bancos de dados. O título de cada sessão apresenta a análise realizada em questão. As páginas criadas e descritas anteriormente permitiram uma análise mais fácil de ser realizada a medida que os dados foram exportados diretamente aos programas desejados.

4.3 BANCO DVMO DO LIHC

Banco com (n=2418) indivíduos, sendo 60,83% mulheres e 39,16% homens.

TABELA 1. Frequência HLA-A DVMO

LOCUS	FREQUENCIA	TOTAL
HLA-A*02	0.26778	1295
HLA-A*24	0.10525	509
HLA-A*01	0.10484	507
HLA-A*03	0.09305	450
HLA-A*68	0.05542	268
HLA-A*31	0.05335	258
HLA-A*29	0.04653	225
HLA-A*23	0.04487	217
HLA-A*11	0.04446	215
HLA-A*30	0.04012	194

HLA-A*26	0.03639	176
HLA-A*32	0.03122	151
HLA-A*33	0.02771	134
HLA-A*25	0.01985	96
HLA-A*66	0.00868	42
HLA-A*74	0.00806	39
HLA-A*34	0.00517	25
HLA-A*36	0.00331	16
HLA-A*69	0.00331	16
HLA-A*80	0.00062	3
Total	1	4836

A análise do gene HLA-A apresentou 20 alelos distintos conforme **TABELA 1**. Teve dominância de HLA-A*02 e frequência de 26%. A soma dos 4 primeiros alelos HLA-A*02, HLA-A*24, HLA-A*01 resulta em 57% da frequência da amostra. A amostra encontra-se dentro do equilíbrio de ligação de HW, com p-valor= 0,0598 para o algoritmo de Guo e Thompson, bem como para cada alelo, frequência observada e esperada que estavam dentro da normalidade.

A análise do gene HLA-B apresentou a frequência descrita na **TABELA 2** abaixo:

TABELA 2. Frequência HLA-B DVMO

LOCUS	FREQUENCIA	TOTAL
HLA-B*35	0.11476	555
HLA-B*44	0.10256	496
HLA-B*51	0.09409	455
HLA-B*15	0.09160	443
HLA-B*07	0.06948	336
HLA-B*08	0.05790	280
HLA-B*14	0.05314	257
HLA-B*18	0.04921	238
HLA-B*40	0.04342	210
HLA-B*39	0.03701	179

HLA-B*49	0.03019	146
HLA-B*38	0.02854	138
HLA-B*57	0.02833	137
HLA-B*27	0.02605	126
HLA-B*58	0.02233	108
HLA-B*50	0.01964	95
HLA-B*52	0.01964	95
HLA-B*53	0.01737	84
HLA-B*13	0.01675	81
HLA-B*45	0.01406	68
HLA-B*55	0.01406	68
HLA-B*41	0.01137	55
HLA-B*37	0.00910	44
HLA-B*42	0.00827	40
HLA-B*48	0.00806	39
HLA-B*56	0.00434	21
HLA-B*47	0.00372	18
HLA-B*81	0.00207	10
HLA-B*73	0.00083	4
HLA-B*54	0.00062	3
HLA-B*46	0.00041	2
HLA-B*78	0.00041	2
HLA-B*59	0.00021	1
HLA-B*67	0.00021	1
HLA-B*82	0.00021	1
Total	1	4836

O grupo alélico com maior frequência foi o HLA-B*35, com pouca diferenciação para as frequências observadas para o HLA-B*44, HLA-B*51 e HLA-B*15. O p-valor da amostra foi 0.8933. As análises de equilíbrio com o algoritmo de Guo e Thompson foram realizadas com um total de 1000000 passos, 2000 passos de desmemorização e 1000 o numero de amostras e tamanho da cadeia de Markov. Diferentemente do HLA-A, o B apresentou 35 diferentes alelos.

Para o HLA-DRB1, as maiores frequências encontradas foram as descritas na **TABELA 3** abaixo:

TABELA 3. Frequência HLA-DRB1 DVMO

LOCUS	FREQUÊNCIA	TOTAL
HLA-DRB1*13	0.13710	663
HLA-DRB1*7	0.12696	614
HLA-DRB1*11	0.12407	600
HLA-DRB1*4	0.12076	584
HLA-DRB1*1	0.10505	508
HLA-DRB1*3	0.10008	484
HLA-DRB1*15	0.08830	427
HLA-DRB1*8	0.07113	344
HLA-DRB1*14	0.03991	193
HLA-DRB1*16	0.03825	185
HLA-DRB1*9	0.01716	83
HLA-DRB1*10	0.01696	82
HLA-DRB1*12	0.01427	69
Total	100.000	4836

NOTA: Foram utilizados somente os dois primeiros caracteres em virtude de mistura de alta, baixa e média resolução dentro do mesmo grupo. O mesmo ocorreu para frequência haplotípica.

Para o caso do antígeno DRB1 foram encontrados 13 alelos distintos com predominância do HLA-DRB1*13 e frequência de 13%. A distribuição dos primeiros alelos está bem equilibrada. O valor de p valor foi de 0.6239 demonstrando-se assim que as frequências encontradas estão respeitando o equilíbrio Hardy-Weinberg.

A análise haplotípica demonstrou que os haplótipos mais comuns foram HLA(A*01B*08DRB1*03) e em segundo HLA(A*29B*44DRB1*07). Interessante observar é que a frequência do grupo alélico não correspondente necessariamente com a frequência haplotípica.

TABELA 4. Frequência haplotípica HLA-A/B/DRB1

HLA-A/HLA-B/HLA-DRB1	Frequência	Total
01:08:03	0.03125	151.1
29:44:07	0.01171	56.6
03:07:15	0.01133	54.8

02:51:13	0.00979	47.4
02:51:08	0.00883	42.7
02:15:13	0.00832	40.2
02:51:11	0.00798	38.6
23:44:07	0.00757	36.6
33:14:01	0.00753	36.4
11:35:01	0.00734	35.5
02:07:15	0.00727	35.1

NOTA: Estão demonstrados somente os resultados de maior frequência.

Ao analisarmos o grupo alélico HLA-A, observamos que o grupo de etnia indígena não obteve tanta variação como os demais grupos étnicos, esta apresentou somente 9 distintos alelos, enquanto as demais apresentaram uma média de 17 alelos distintos. A etnia negra não apresentou equilíbrio para os homozigotos, apresentando p-valor=0,0308.

TABELA 5. Frequência HLA-A - DVMO por etnias

HLA-A	Pardo		Não Informado		Negro		Índio		Branco		Asiático	
	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f
1	0.08422	63	0.18605	16	0.08108	18	0.08333	1	0.10886	403	0.09091	6
2	0.26604	199	0.30233	26	0.21622	48	0.16667	2	0.27093	1003	0.25758	17
3	0.10027	75	0.08140	7	0.04505	10	*	*	0.09562	354	0.06061	4
11	0.03342	25	0.02326	2	0.02703	6	0.08333	1	0.04781	177	0.06061	4
23	0.05481	41	0.03488	3	0.11261	25	*	*	0.03944	146	0.03030	2
24	0.10963	82	0.11628	10	0.08559	19	0.08333	1	0.10454	387	0.15152	10
25	0.01604	12	0.01163	1	0.00450	1	*	*	0.02188	81	0.01515	1
26	0.02807	21	0.03488	3	0.02252	5	0.16667	2	0.03674	136	0.13636	9
29	0.03342	25	0.03488	3	0.05856	13	*	*	0.04943	183	0.01515	1
30	0.05214	39	0.03488	3	0.06757	15	0.08333	1	0.03647	135	0.01515	1
31	0.06150	46	0.06977	6	0.06306	14	0.08333	1	0.04997	185	0.09091	6
32	0.02674	20	0.01163	1	0.01802	4	*	*	0.03404	126	0.03030	2
33	0.02406	18	0.01163	1	0.03153	7	0.08333	1	0.02836	105	*	*
34	0.00802	6	*	*	0.01802	4	*	*	0.00405	15	*	*
36	0.00936	7	*	*	0.00450	1	*	*	0.00216	8	*	*
66	0.00936	7	0.01163	1	0.02252	5	*	*	0.00783	29	*	*
68	0.06551	49	0.02326	2	0.07207	16	0.16667	2	0.05294	196	0.04545	3
69	0.00267	2	0.01163	1	*	*	*	*	0.00351	13	*	*
74	0.01471	11	*	*	0.04505	10	*	*	0.00486	18	*	*
80	*	*	*	*	0.00450	1	*	*	0.00054	2	*	*
Total		748		86		222		12		3702		66

NOTA: " * " Representam frequências não encontradas.

É observável que há um consenso entre as etnias onde o HLA-A*02 é o mais freqüente para todas as estudadas. Porém para a etnia parda a segunda frequência mais observada trata-se da HLA-A*24 enquanto para a dos não informados a segunda mais observada é a HLA-A*01, a mesma que se repete para a branca. Já nos negros há que se destaca é a HLA-A*23.

Para o alelo HLA-B as frequências encontradas já são um pouco diferentes. Onde a frequência HLA-B*35 é mais comum apenas para a etnia parda, branca e asiática. Nas demais a frequência oscila, no caso dos negros HLA-B*15, para os não identificados HLA-B*51 e para os indígenas HLA-B*15 e 38, porém nesse último somente foram encontrados dois indivíduos com a mesma tipagem. A diversidade é maior na etnia branca e na parda, valores que podem ser explicados em virtude da maior quantidade de indivíduos analisados. A **TABELA 6** demonstra os resultados encontrados divididos nas etnias estudadas. Todas as amostras apresentaram p.valor dentro do esperado, estando assim dentro do equilíbrio de Hardy-Weinberg.

TABELA 6. Frequência HLA-B DVMO por etnias

HLA-B	Pardo		Não identificado		Negro		Indio		Branco		Asiático	
	f	frel	f	frel	f	frel	f	frel	f	frel	f	frel
7	0.05749	43	0.06977	6	0.09910	22	*	*	0.07023	260	0.07576	5
8	0.04545	34	0.10465	9	0.06306	14	*	*	0.05943	220	0.04545	3
13	0.00936	7	0.02326	2	0.01351	3	*	*	0.01864	69		
14	0.05080	38	0.03488	3	0.05405	12	0.08333	1	0.05457	202	0.01515	1
15	0.11364	85	0.06977	6	0.15315	34	0.16667	2	0.08347	309	0.10606	7
18	0.04545	34	0.01163	1	0.04054	9	0.08333	1	0.05159	191	0.03030	2
27	0.02540	19	0.05814	5	0.01351	3	*	*	0.02593	96	0.04545	3
35	0.13636	102	0.10465	9	0.05856	13	0.08333	1	0.11426	423	0.10606	7
37	0.00535	4	0.01163	1	0.00901	2	*	*	0.00972	36	0.01515	1
38	0.02139	16	0.02326	2	0.01802	4	0.16667	2	0.03025	112	0.03030	2
39	0.04011	30	0.02326	2	0.04505	10	0.08333	1	0.03620	134	0.03030	2
40	0.04412	33	0.04651	4	0.03604	8	*	*	0.04295	159	0.09091	6
41	0.01070	8	*		0.01351	3	*	*	0.01162	43	0.01515	1
42	0.01070	8	0.01163	1	0.04054	9	*	*	0.00567	21	0.01515	1
44	0.09626	72	0.09302	8	0.08108	18	0.08333	1	0.10562	391	0.09091	6
45	0.01070	8	0.01163	1	0.01802	4	0.08333	1	0.01459	54	*	*
46	*	*	*	*	*	*	*	*	0.00027	1	0.01515	1
47	0.00668	5	*	*	0.00450	1	*	*	0.00324	12	*	*
48	0.01203	9	0.02326	2	0.00450	1	*	*	0.00729	27	*	*
49	0.02941	22	0.02326	2	0.03153	7	*	*	0.03079	114	0.01515	1
50	0.02273	17	0.02326	2	0.02252	5	*	*	0.01918	71	*	*
51	0.08021	60	0.12791	11	0.03604	8	0.08333	1	0.09941	368	0.10606	7
52	0.02139	16	0.02326	2	0.00901	2	*	*	0.01945	72	0.04545	3
53	0.02807	21	0.01163	1	0.05856	13	*	*	0.01324	49	*	*
54	*	*	*	*	*	*	*	*	*	*	0.04545	3
55	0.00936	7	0.01163	1	0.01802	4	*	*	0.01486	55	0.01515	1
56	0.00535	4	0.01163	1	*	*	*	*	0.00432	16	*	*
57	0.02941	22	0.03488	3	0.02703	6	0.08333	1	0.02836	105	*	*
58	0.02540	19	0.01163	1	0.03153	7	*	*	0.02161	80	0.01515	1
59	*	*	*	*	*	*	*	*	*	*	0.01515	1
67	*	*	*	*	*	*	*	*	0.00108	4	0.01515	1
78	*	*	*	*	*	*	*	*	0.00054	2	*	*
81	0.00668	5	*	*	*	*	*	*	0.00135	5	*	*
Total		748		86		222		12		3702		66

NOTA: " * " Representam frequências não encontradas.

O grupo alélico DRB1, mostrou que a maior frequência para as etnias pardo, branco e negro foi HLA-DRB1*13, porém para os indígenas e asiático as frequências maiores são outras. Para o grupo dos asiáticos a frequência mais comum foi HLA-DRB1*15, enquanto para os indígenas a de maior frequência foi a HLA-DRB1*4. Curiosamente as que obtiveram a segunda maior frequência são diferentes entre os grupos. Para os pardos a segunda maior é a HLA-DRB1*04, seguida pela 11, já para os negros e brancos a 7, para os indígenas um empate entre 13, 11 e 3, para os asiáticos a 4 assim como os pardos. Para nenhuma das etnias foi encontrado p.valor inferior a 0,05. Ou seja todos encontram-se em equilíbrio. A **TABELA 7** demonstra os resultados.

TABELA 7. Frequência HLA-DRB1 DVMO por etnias

HLA-DR	Pardo		Não identificado		Negro		Indio		Branco		Asiático	
	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f
1	0.10160	76	0.11628	10	0.09009	20	0.08333	1	0.10697	396	0.07576	5
3	0.10294	77	0.11628	10	0.11261	25	0.16667	2	0.09860	365	0.07576	5
4	0.12166	91	0.11628	10	0.10360	23	0.25000	3	0.12021	445	0.18182	12
7	0.10561	79	0.10465	9	0.13964	31	0.08333	1	0.13236	490	0.06061	4
8	0.08824	66	0.09302	8	0.04505	10	*	*	0.06888	255	0.07576	5
9	0.02807	21	0.02326	2	0.01351	3	*	*	0.01432	53	0.06061	4
10	0.01872	14	*	*	0.02703	6	*	*	0.01648	61	0.01515	1
11	0.12032	90	0.12791	11	0.10360	23	0.16667	2	0.12696	470	0.06061	4
12	0.01738	13	0.01163	1	0.01802	4	*	*	0.01297	48	0.04545	3
13	0.13503	101	0.16279	14	0.14414	32	0.16667	2	0.13749	509	0.07576	5
14	0.05214	39	0.02326	2	0.03604	8	*	*	0.03782	140	0.06061	4
15	0.07086	53	0.06977	6	0.11712	26	0.08333	1	0.08860	328	0.19697	13
16	0.03743	28	0.03488	3	0.04955	11	*	*	0.03836	142	0.01515	1
Total		748		86		222		12		3702		66

NOTA: " * " Representam frequências não encontradas.

Só foram considerados os 2 primeiros caracteres, isso em virtude dos dados estarem a média resolução e pouquíssimos a alta resolução. Correspondente a uma tipagem de baixa resolução.

4.4 BANCO DA IMUNOGENÉTICA DO LIHC (TOTAL DE INDIVÍDUOS SELECIONADOS – UM DE CADA FAMÍLIA OU OS PAIS QUANDO NÃO CONSANGUÍNEOS)

Esta análise foi realizada através da seleção aleatória de somente 1 indivíduo de cada família, eliminando os familiares consanguíneos da amostra. A seleção desses indivíduos foi realizada por meio de código SQL com a função DISTINCT e algumas condições. Diferentemente da análise anterior esta foi realizada com tipagem a alta resolução, onde o resultado é mais preciso e envolve 4 caracteres.

A análise do alelo HLA-A demonstra que esse grupo obteve o resultado de HLA-A*0201 como o mais frequente, somando 64 indivíduos e uma frequência de 21%. Em seguida temos a frequência HLA-A*0101 com 9%, com um total de 28 indivíduos. A amostra também está dentro do equilíbrio Hardy-Weinberg (HW), com $p\text{-valor}=0,2621$. Todas as demais frequências podem ser observadas na tabela abaixo.

TABELA 8. Frequência HLA-A Amostra 2

LOCUS	Frequência	Total
HLA-A*0201	0.21769	64
HLA-A*0101	0.09524	28
HLA-A*2402	0.08503	25
HLA-A*0301	0.07823	23
HLA-A*2902	0.06803	20
HLA-A*3101	0.06463	19
HLA-A*1101	0.06122	18
HLA-A*6801	0.05442	16
HLA-A*3201	0.03741	11
HLA-A*2301	0.03061	9
HLA-A*2601	0.02381	7
HLA-A*3001	0.02041	6
HLA-A*3002	0.02041	6
HLA-A*3301	0.02041	6
HLA-A*0205	0.01701	5
HLA-A*2501	0.01361	4
HLA-A*2403	0.01020	3
HLA-A*0202	0.00680	2
HLA-A*0204	0.00680	2
HLA-A*0302	0.00680	2
HLA-A*3004	0.00680	2
HLA-A*3303	0.00680	2

HLA-A*7401	0.00680	2
HLA-A*0102	0.00340	1
HLA-A*0206	0.00340	1
HLA-A*0211	0.00340	1
HLA-A*2401	0.00340	1
HLA-A*2901	0.00340	1
HLA-A*3115	0.00340	1
HLA-A*3402	0.00340	1
HLA-A*3601	0.00340	1
HLA-A*6601	0.00340	1
HLA-A*6802	0.00340	1
HLA-A*6901	0.00340	1
HLA-A*8001	0.00340	1
Total	1	294

Já a análise do HLA-B demonstrou que a maior frequência foi a de HLA-B*0702, com 7%, seguida por HLA-B*4403 e HLA-B*0801. As frequências, pelo menos as 5 primeiras, demonstraram-se muito próximas. Isso demonstra uma harmonia na população estudada. Esse resultado apresentou p.valor dentro do equilíbrio de HW. A **TABELA 9** apresenta todos os resultados obtidos.

TABELA 9. Frequência HLA-B Amostra 2

LOCUS	Frequência	Total
HLA-B*0702	0.07031	27
HLA-B*4403	0.06771	26
HLA-B*0801	0.06510	25
HLA-B*1402	0.06250	24
HLA-B*3501	0.05729	22
HLA-B*5101	0.03906	15
HLA-B*1501	0.03646	14
HLA-B*4002	0.03385	13
HLA-B*1801	0.02865	11
HLA-B*4901	0.02865	11
HLA-B*1503	0.02604	10
HLA-B*5301	0.02604	10
HLA-B*3503	0.02344	9
HLA-B*4001	0.02344	9
HLA-B*4402	0.02083	8
HLA-B*5201	0.02083	8
HLA-B*5701	0.02083	8
HLA-B*1302	0.01823	7
HLA-B*4501	0.01823	7
HLA-B*5801	0.01823	7
HLA-B*1401	0.01562	6
HLA-B*1517	0.01562	6

HLA-B*4201	0.01562	6
HLA-B*5001	0.01562	6
HLA-B*3502	0.01302	5
HLA-B*3801	0.01302	5
HLA-B*4004	0.01302	5
HLA-B*2705	0.01042	4
HLA-B*3505	0.01042	4
HLA-B*3508	0.01042	4
HLA-B*4006	0.01042	4
HLA-B*5501	0.01042	4
HLA-B**140	0.00781	3
HLA-B*1504	0.00781	3
HLA-B*3701	0.00781	3
HLA-B*3905	0.00781	3
HLA-B*4102	0.00781	3
HLA-B*4802	0.00781	3
HLA-B*5703	0.00781	3
HLA-B*0705	0.00521	2
HLA-B*1510	0.00521	2
HLA-B*1516	0.00521	2
HLA-B*2702	0.00521	2
HLA-B*3504	0.00521	2
HLA-B*3901	0.00521	2
HLA-B*4701	0.00521	2
HLA-B**150	0.00260	1
HLA-B**151	0.00260	1
HLA-B*1518	0.00260	1
HLA-B*1520	0.00260	1
HLA-B*2707	0.00260	1
HLA-B*3510	0.00260	1
HLA-B*3516	0.00260	1
HLA-B*3902	0.00260	1
HLA-B*3909	0.00260	1
HLA-B*4011	0.00260	1
HLA-B*4043	0.00260	1
HLA-B*4202	0.00260	1
HLA-B*4405	0.00260	1
HLA-B*4801	0.00260	1
HLA-B*5108	0.00260	1
HLA-B*5401	0.00260	1
HLA-B*5601	0.00260	1
HLA-B*5702	0.00260	1
HLA-B*5802	0.00260	1
Total	1	384

As frequências analisadas para o antígeno HLA-DRB1 demonstraram que o com maior frequência foi o HLA-DR*0701 com 13% seguido pelo HLA-DRB1*1501

com 138 indivíduos e 9%. O p.valor dessa amostragem foi igual a 0.1620, demonstrando assim dentro do esperado de HW. Resultados podem ser conferidos na **TABELA 10**.

TABELA 10. Frequência HLA-DRB1 Amostra 2

LOCUS	Frequência	Total
HLA-DRB1*0701	0.13479	203
HLA-DRB1*1501	0.09163	138
HLA-DRB1*0301	0.08699	131
HLA-DRB1*1101	0.06574	99
HLA-DRB1*0101	0.06308	95
HLA-DRB1*1301	0.06175	93
HLA-DRB1*1302	0.03918	59
HLA-DRB1*1104	0.03187	48
HLA-DRB1*0102	0.02988	45
HLA-DRB1*1401	0.02922	44
HLA-DRB1*0404	0.02722	41
HLA-DRB1*0801	0.02590	39
HLA-DRB1*0401	0.02457	37
HLA-DRB1*0402	0.02125	32
HLA-DRB1*1602	0.02058	31
HLA-DRB1*1503	0.01992	30
HLA-DRB1*0405	0.01726	26
HLA-DRB1*1001	0.01726	26
HLA-DRB1*0302	0.01660	25
HLA-DRB1*1201	0.01660	25
HLA-DRB1*0411	0.01461	22
HLA-DRB1*0403	0.01328	20
HLA-DRB1*0901	0.01328	20
HLA-DRB1*1102	0.01328	20
HLA-DRB1*0802	0.01195	18
HLA-DRB1*1303	0.01195	18
HLA-DRB1*1601	0.01129	17
HLA-DRB1*0407	0.00996	15
HLA-DRB1*0804	0.00996	15
HLA-DRB1*0103	0.00797	12
HLA-DRB1*1103	0.00730	11
HLA-DRB1*1502	0.00664	10
HLA-DRB1*1402	0.00465	7
HLA-DRB1*0807	0.00398	6
HLA-DRB1*0104	0.00332	5
HLA-DRB1*0408	0.00332	5
HLA-DRB1*0803	0.00266	4
HLA-DRB1*1406	0.00199	3
HLA-DRB1*1305	0.00133	2
HLA-DRB1*0406	0.00066	1

HLA-DRB1*0409	0.00066	1
HLA-DRB1*1202	0.00066	1
HLA-DRB1*1203	0.00066	1
HLA-DRB1*1331	0.00066	1
HLA-DRB1*1405	0.00066	1
HLA-DRB1*1413	0.00066	1
HLA-DRB1*1603	0.00066	1
HLA-DRB1*40551	0.00066	1
Total	1	1506

A frequência de haplótipos resultou na **TABELA 11**, onde obteve-se o haplótipo HLA-A*B*DRB1(0101:0801:0301) com 11 indivíduos seguido por HLA-A*B*DRB1(2902:4403:0701).

TABELA 11. Frequência haplotípica HLA-A*B*DRB1 Amostra 2

HLA(A*B*DRB1)	Frequência	Total
0101:0801:0301:	0.04583	11.0
2902:4403:0701:	0.04167	10.0
0301:0702:1501:	0.02500	6.0
0301:0801:0301:	0.01250	3.0
2301:4403:0701:	0.01250	3.0
0201:3501:1501:	0.01250	3.0
0301:3501:0101:	0.01250	3.0
0201:1801:1301:	0.01250	3.0
3002:1402:1503:	0.01250	3.0
0201:4001:0404:	0.01250	3.0
0201:0702:0101:	0.01250	3.0
1101:3501:1104:	0.01250	3.0
0101:5701:0701:	0.01150	2.8
0201:4403:0701:	0.00833	2.0

4.5 BANCO DA IMUNOGENÉTICA DO HC (FAMILIARES DE PACIENTES SEM AAS E AF)

TABELA 12. Frequência HLA-A Amostra 3

LOCUS	Frequência	Total
HLA-A*0201	0.22549	69
HLA-A*0101	0.09804	30
HLA-A*2402	0.08170	25
HLA-A*0301	0.07516	23
HLA-A*2902	0.06863	21

HLA-A*3101	0.06209	19
HLA-A*1101	0.05882	18
HLA-A*6801	0.05882	18
HLA-A*3201	0.03595	11
HLA-A*2301	0.03268	10
HLA-A*2601	0.02288	7
HLA-A*3001	0.01961	6
HLA-A*3002	0.01961	6
HLA-A*3301	0.01961	6
HLA-A*0205	0.01634	5
HLA-A*2501	0.01307	4
HLA-A*2403	0.00980	3
HLA-A*0202	0.00654	2
HLA-A*0204	0.00654	2
HLA-A*0302	0.00654	2
HLA-A*3004	0.00654	2
HLA-A*3303	0.00654	2
HLA-A*6802	0.00654	2
HLA-A*7401	0.00654	2
HLA-A*0102	0.00327	1
HLA-A*0206	0.00327	1
HLA-A*0211	0.00327	1
HLA-A*2401	0.00327	1
HLA-A*2901	0.00327	1
HLA-A*3115	0.00327	1
HLA-A*3402	0.00327	1
HLA-A*3601	0.00327	1
HLA-A*6601	0.00327	1
HLA-A*6901	0.00327	1
HLA-A*8001	0.00327	1
Total	1	306

A análise do p.valor por Guo e Thompson resultou em desequilíbrio. O p.valor neste algoritmo teve o valor de 0.0076, porém para a análise original do equilíbrio de HW com fator de corte 5, porém sem realizar nenhum corte o programa PyPop, afirma que a população encontra-se em equilíbrio tanto para heterozigotos como homozigotos. Nesse caso p.valor= 0.9872 para heterozigotos e 0.9320 para homozigóticos. O mesmo aconteceu para a frequência obtida para o antígeno HLA-DRB1, porém com o diferencial de p.valor= 0.0034** para o algoritmo de Guo e Thompson e p.valor acima de 0,05 somente para análise com corte ou heterozigóticos.

A análise revelou maior frequência do antígeno HLA-A*0201 com valor expressivo de 22% da amostra já a análise da do HLA-B revelou uma hegemonia entre os dados onde HLA-B*0801, HLA-B*0702, HLA-B*4403 e HLA-B*1402 com variação entre 6 e 7% da amostragem. Os dados podem ser conferidos na **TABELA 12** e **TABELA 13**.

TABELA 13. Frequência HLA-B Amostra 3

LOCUS	Frequência	Total
HLA-B*0801	0.07071	28
HLA-B*0702	0.06818	27
HLA-B*4403	0.06818	27
HLA-B*1402	0.06061	24
HLA-B*3501	0.05556	22
HLA-B*5101	0.03788	15
HLA-B*1501	0.03535	14
HLA-B*4002	0.03283	13
HLA-B*4901	0.03030	12
HLA-B*1801	0.02778	11
HLA-B*5301	0.02778	11
HLA-B*1503	0.02525	10
HLA-B*4001	0.02525	10
HLA-B*3503	0.02273	9
HLA-B*5701	0.02273	9
HLA-B*1302	0.02020	8
HLA-B*4402	0.02020	8
HLA-B*4501	0.02020	8
HLA-B*5201	0.02020	8
HLA-B*5801	0.01768	7
HLA-B*1401	0.01515	6
HLA-B*1517	0.01515	6
HLA-B*4004	0.01515	6
HLA-B*4201	0.01515	6
HLA-B*5001	0.01515	6
HLA-B*3502	0.01263	5
HLA-B*3801	0.01263	5
HLA-B*2705	0.01010	4
HLA-B*3505	0.01010	4
HLA-B*3508	0.01010	4
HLA-B*4006	0.01010	4
HLA-B*5501	0.01010	4
HLA-B**140	0.00758	3
HLA-B*1504	0.00758	3
HLA-B*3701	0.00758	3
HLA-B*3905	0.00758	3
HLA-B*4102	0.00758	3

HLA-B*4802	0.00758	3
HLA-B*5703	0.00758	3
HLA-B*0705	0.00505	2
HLA-B*1510	0.00505	2
HLA-B*1516	0.00505	2
HLA-B*1518	0.00505	2
HLA-B*2702	0.00505	2
HLA-B*3504	0.00505	2
HLA-B*3901	0.00505	2
HLA-B*4701	0.00505	2
HLA-B**150	0.00253	1
HLA-B**151	0.00253	1
HLA-B*1520	0.00253	1
HLA-B*2707	0.00253	1
HLA-B*3510	0.00253	1
HLA-B*3516	0.00253	1
HLA-B*3902	0.00253	1
HLA-B*3909	0.00253	1
HLA-B*4011	0.00253	1
HLA-B*4043	0.00253	1
HLA-B*4202	0.00253	1
HLA-B*4405	0.00253	1
HLA-B*4801	0.00253	1
HLA-B*5108	0.00253	1
HLA-B*5401	0.00253	1
HLA-B*5601	0.00253	1
HLA-B*5702	0.00253	1
HLA-B*5802	0.00253	1
Total	1	396

A **TABELA 14** mostra as frequências obtidas onde a de maior destaque foi a 0701 com 13% seguido pela 1501 com 8%. Teve p.valor dentro do aceitável somente para os heterozigóticos. A análise através do algoritmo de Guo e Thompson como dito anteriormente não demonstrou equilíbrio de HW.

TABELA 14. Frequência HLA-DRB1 Amostra 3

LOCUS	Frequência	Total
HLA-DRB1*0701	0.13648	220
HLA-DRB1*1501	0.08995	145
HLA-DRB1*0301	0.08747	141
HLA-DRB1*1101	0.06638	107
HLA-DRB1*1301	0.06514	105
HLA-DRB1*0101	0.06266	101

HLA-DRB1*1302	0.03846	62
HLA-DRB1*1104	0.03226	52
HLA-DRB1*1401	0.03040	49
HLA-DRB1*0102	0.02792	45
HLA-DRB1*0801	0.02792	45
HLA-DRB1*0404	0.02730	44
HLA-DRB1*0401	0.02357	38
HLA-DRB1*0402	0.02109	34
HLA-DRB1*1602	0.02047	33
HLA-DRB1*1001	0.01861	30
HLA-DRB1*1503	0.01861	30
HLA-DRB1*1201	0.01737	28
HLA-DRB1*0405	0.01675	27
HLA-DRB1*0302	0.01551	25
HLA-DRB1*0403	0.01489	24
HLA-DRB1*0411	0.01365	22
HLA-DRB1*0802	0.01303	21
HLA-DRB1*0901	0.01303	21
HLA-DRB1*1102	0.01241	20
HLA-DRB1*1303	0.01117	18
HLA-DRB1*1601	0.01055	17
HLA-DRB1*0407	0.00931	15
HLA-DRB1*0804	0.00931	15
HLA-DRB1*0103	0.00868	14
HLA-DRB1*1103	0.00682	11
HLA-DRB1*1502	0.00620	10
HLA-DRB1*0807	0.00496	8
HLA-DRB1*1402	0.00434	7
HLA-DRB1*0104	0.00310	5
HLA-DRB1*0408	0.00310	5
HLA-DRB1*0803	0.00248	4
HLA-DRB1*1406	0.00186	3
HLA-DRB1*1305	0.00124	2
HLA-DRB1*0406	0.00062	1
HLA-DRB1*0409	0.00062	1
HLA-DRB1*1202	0.00062	1
HLA-DRB1*1203	0.00062	1
HLA-DRB1*1331	0.00062	1
HLA-DRB1*1405	0.00062	1
HLA-DRB1*1413	0.00062	1
HLA-DRB1*1603	0.00062	1
HLA-DRB1*40551	0.00062	1
Total	1	1612

As frequências haplotípicas de maior destaque foram HLA-A/HLA-B/HLA-DRB1(0101:0801:0301) com apenas 5%, valor bem pouco expressivo para a amostragem como mostra a

TABELA 15. Teve 119 fenótipos únicos, com 455 genótipos e o total de haplótipos foi de 778.

TABELA 15. Frequência haplotípica HLA-A*B*DRB1 Amostra 3

HLA-A/HLA-B/HLA-DRB1	Frequência	Total
0101:0801:0301:	0.05242	13.0
2902:4403:0701:	0.04032	10.0
0301:0702:1501:	0.02419	6.0
2301:4403:0701:	0.01613	4.0
0201:4001:0404:	0.01613	4.0
0301:0801:0301:	0.01210	3.0
0201:0702:0101:	0.01210	3.0
0201:1801:1301:	0.01210	3.0
3002:1402:1503:	0.01210	3.0
0201:3501:1501:	0.01210	3.0
0101:5701:0701:	0.01113	2.8
1101:3501:1104:	0.00806	2.0
0301:3501:1601:	0.00806	2.0
3101:3503:0101:	0.00806	2.0
0201:4402:1101:	0.00806	2.0
2402:3502:1104:	0.00806	2.0
0201:1402:1101:	0.00806	2.0
6802:5301:1302:	0.00806	2.0
2402:5201:1502:	0.00806	2.0
3101:3905:0701:	0.00806	2.0
6801:4004:0807:	0.00806	2.0
3301:1402:0102:	0.00806	2.0
0201:4501:0405:	0.00806	2.0
2902:1518:1301:	0.00806	2.0
0101:0801:0701:	0.00806	2.0
0201:3501:1101:	0.00806	2.0
0201:4403:0701:	0.00806	2.0
2402:1402:0301:	0.00806	2.0
3001:4201:0302:	0.00806	2.0
3201:4002:1501:	0.00806	2.0
3001:1302:0701:	0.00806	2.0
0201:4501:0403:	0.00806	2.0
1101:3501:0801:	0.00806	2.0

A análise da tipagem HLA-A para a amostragem 3, porém em separado para cada etnia revelou-se dentro do equilíbrio de HW. Porém para a tipagem B e DRB1, a etnia branca encontrou-se fora do equilíbrio, com p.valor 0,0117 e 0,0336 respectivamente. Na

TABELA 16, pode-se observar uma dominância do antígeno HLA-A*0201 o que acontece também para o grupo dos que não informaram a etnia no cadastro. Fato que o segundo grupo de maior frequência para a população estudada trata-se da tipagem 0101 seguida pela 2901 e 2402, valores estes que não obtiveram tanta expressão para a etnia branca. A qual o segundo valor de maior frequência foi HLA-A*0301.

TABELA 16. Frequência HLA-A Amostra 3 por etnia

HLA-A	Branco		Negro		Mulato		Índio		Cafuzo		Não Informado		Indiano		Pardo	
	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f
0101	0.02778	1			0.25000	1					0.10687	28				
0102											0.00382	1				
0201	0.30556	11			0.25000	1					0.21374	56			0.50000	1
0202					0.25000	1					0.00763	2				
0204											0.00382	1				
0205	0.02778	1									0.01527	4				
0206											0.00382	1				
0211											0.00382	1				
0301	0.19444	7									0.06107	16				
0302											0.00763	2				
1101	0.05556	2									0.06107	16				
2301	0.05556	2									0.03053	8				
2401											0.00382	1				
2402	0.08333	3			0.25000	1					0.08015	21				
2403	0.02778	1									0.00763	2				
2501	0.02778	1									0.01145	3				
2601	0.02778	1									0.02290	6				
2901											0.00382	1				
2902											0.08015	21				
3001	0.02778	1									0.01908	5				
3002											0.02290	6				
3303			0.50000	1												
3004											0.00763	2				
3101	0.08333	3									0.05725	15			0.50000	1
3115											0.00382	1				

3201				0.04198	11	
3301				0.02290	6	
3303				0.00382	1	
3402				0.00382	1	
3601				0.00382	1	
6601				0.00382	1	
6801	0.05556	2		0.06107	16	
6802				0.00763	2	
6901				0.00382	1	
7401				0.00763	2	
8001		0.50000	1			
Total	36	2	4		262	2

A frequência observada de maior valor foi a HLA-B*3501 e 0801 para a etnia branca e para o grupo de etnia não informada. Fator interessante é que o grupo dos que não informaram a etnia apresenta destaque também para a tipagem 0702, 1402 e 4403, um forte indício de que neste grupo não está presente somente a etnia branca. Para a amostragem da **TABELA 17** e **TABELA 18** a etnia branca revelou-se estar fora do equilíbrio de HW. No antígeno DRB1, valores muito próximos entre o grupo de brancos e de não informados, obtendo destaque a HLA-DRB1*0301, 0701, 1101 e 0101.

TABELA 17. Frequência HLA-B Amostra 3 por etnia

HLA-B	Branco		Negro		Mulato		Índio		Cafuzo		Não Informado		Indiano		Pardo	
	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f
0702	0.04348	2									0.07576	25				
0705											0.00606	2				
0801	0.08696	4									0.07273	24				
1302	0.02174	1									0.02121	7				
1401											0.01818	6				
1402	0.02174	1	0.16667	1	0.20000	2					0.06061	20				
1501	0.02174	1			0.10000	1					0.03636	12				
1503	0.02174	1			0.10000	1					0.02424	8				
1504							0.25000	2			0.00909	3				
1510											0.00606	2				
1516											0.00606	2				
1517	0.02174	1			0.10000	1					0.01212	4				
1518											0.00606	2				
1520											0.00303	1				
1801	0.02174	1									0.03030	10				
2702											0.00606	2				
2705											0.01212	4				
2707											0.00303	1				
3501	0.08696	4	0.16667	1							0.05152	17				
3502	0.02174	1			0.10000	1					0.00909	3				
3503	0.04348	2									0.02121	7				
3504											0.00606	2				
3505							0.62500	5			0.01212	4				
3508	0.02174	1									0.00909	3				
3510	0.02174	1														
3516											0.00303	1				
3701											0.00909	3				
3801											0.01515	5				

3901								0.00606	2
3902								0.00303	1
3905								0.00909	3
3909								0.00303	1
4001								0.03030	10
4002	0.02174	1						0.03636	12
4004	0.02174	1						0.01515	5
4006								100.000	4
4011								0.00303	1
4043								0.00303	1
4102	0.02174	1						0.00606	2
4201			0.16667	1				0.01515	5
4402	0.02174	1	0.16667	1				0.02121	7
4403	0.04348	2						0.07576	25
4405								0.00303	1
4501	0.04348	2						0.01818	6
4701	0.02174	1						0.00303	1
4801								0.00303	1
4802	0.02174	1						0.00606	2
4901								0.03636	12
5001	0.04348	2		0.10000	1			0.00909	3
5101	0.02174	1		0.10000	1			0.03939	13
5108								0.00303	1
5201				0.10000	1	0.12500	1	0.02121	7
5301	0.02174	1	0.33333	2		100.000	8	0.02424	8
5401								0.00303	1
5501	0.04348	2						0.00606	2
5601								0.00303	1
5701	0.06522	3						0.01818	6
5702	0.02174	1							
5703	0.02174	1		0.10000	1			0.00303	1

5801	0.04348	2							0.01515	5			
5802									0.00303	1			
	46		6		10					330		4	

TABELA 18. Frequência HLA-DRB1 Amostra 3 por etnia

HLA-	Branco frel	f	Negro frel	f	Mulato frel	f	Índio frel	f	Cafuzo frel	f	Não Informado frel	f	Indiano frel	f	Pardo frel	f
0101	0.06954	21									0.06359	80			0.25000	1
0102	0.03974	12			0.02632	1					0.02464	31				
0103	0.00993	3									0.00874	11				
0104	0.00331	1							0.50000	1	0.00238	3				
0301	0.09603	29			0.13158	5					0.08506	107				
0302	0.00331	1	0.50000	2	0.02632	1					0.01669	21				
0401	0.01325	4									0.02703	34				
0402	0.04305	13			0.02632	1					0.01590	20				
0403	0.01325	4									0.01590	20				
0404	0.03311	10									0.02703	34				
0405	0.00993	3			0.02632	1					0.01828	23				
0406	0.00331	1														
0407	0.00662	2									0.01033	13				
0408	0.00331	1									0.00238	3			0.25000	1
0409	0.00331	1														
0411	0.01987	6			0.05263	2					0.00954	12			0.50000	2
0701	0.11258	34			0.15789	6					0.14308	180				
0801	0.02980	9									0.02862	36				
0802							0.50000	4			0.01351	17				
0803											0.00318	4				
0804	0.00993	3	0.50000	2							0.00795	10				
0807	0.00993	3	100.000	4							0.00397	5				

4.6 BANCO DO LGMH

Para o cálculo das frequências foi utilizado somente o antígeno HLA-DRB1, tal procedimento ocorreu em virtude do banco só conter dados em alta resolução para os HL-DRB1, DQB1 E DQA1. As frequências encontradas para HLA-DRB1 foram as descritas abaixo:

TABELA 19. Frequência HLA-DRB1 LGMH

Antígeno	Frequency	(Count)
HLA-DRB1*0701	0.09692	44
HLA-DRB1*0301	0.07489	34
HLA-DRB1*0101	0.06608	30
HLA-DRB1*1101	0.06608	30
HLA-DRB1*1501	0.06608	30
HLA-DRB1*1302	0.05727	26
HLA-DRB1*1503	0.05727	26
HLA-DRB1*1301	0.04626	21
HLA-DRB1*1401	0.04185	19
HLA-DRB1*1602	0.03524	16
HLA-DRB1*0801	0.03304	15
HLA-DRB1*1104	0.02643	12
HLA-DRB1*0102	0.02423	11
HLA-DRB1*1402	0.02423	11
HLA-DRB1*1001	0.02203	10
HLA-DRB1*1102	0.02203	10
HLA-DRB1*0411	0.01982	9
HLA-DRB1*0802	0.01982	9
HLA-DRB1*1601	0.01762	8
HLA-DRB1*0804	0.01542	7
HLA-DRB1*0302	0.01322	6
HLA-DRB1*0401	0.01322	6
HLA-DRB1*1201	0.01322	6
HLA-DRB1*9012	0.01322	6
HLA-DRB1*0403	0.01101	5
HLA-DRB1*0405	0.01101	5
HLA-DRB1*0402	0.00881	4
HLA-DRB1*1303	0.00881	4
HLA-DRB1*1406	0.00881	4
HLA-DRB1*1502	0.00881	4
HLA-DRB1*0407	0.00661	3
HLA-DRB1*0408	0.00661	3
HLA-DRB1*0807	0.00661	3
HLA-DRB1*1103	0.00661	3
HLA-DRB1*1404	0.00661	3

HLA-DRB1*-0804	0.00441	2
HLA-DRB1*-1302	0.00441	2
HLA-DRB1*0404	0.00441	2
HLA-DRB1*-1101	0.00220	1
HLA-DRB1*-1301	0.00220	1
HLA-DRB1*-1401	0.00220	1
HLA-DRB1*0103	0.00220	1
HLA-DRB1*0803	0.00220	1
Total	1	454

Para esta amostra o antígeno DRB1 mais comum foi o HLA-DRB1*0401, com frequência de 9% seguido pelo 0301 com frequência de 7%. A amostragem encontrou-se dentro do equilíbrio HW, com p.valor=0,3827. Uma análise somente considerando os primeiros 2 dígitos da tipagem DRB1, mostrou que o mais frequente é o HLA-DR*15, com 13% , enquanto em segundo lugar o antígeno HLA-DR*11 aparece com 12%. Para esta análise o p.valor encontrado foi de 0,4826 no algoritmo de Guo e Thompson. A análise das frequências dividida em grupos étnicos pode ser encontrada na TABELA 20:

TABELA 20. Frequência HLA-DRB1 LGMH por etnias

HLA- DRB1	Branco		Mulato Claro Europeu		Mulato Claro Brasileiro		Mulato Escuro		Mulato Medio		Negro	
	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f
0101	0.09174	20	0.05556	7					0.04167	3		
0102	0.00917	2	0.03175	4	0.10000	2	0.10000	2	0.04167	3		
0103			0.00794	1								
0301	0.08257	18	0.07937	10					0.05556	4	0.11111	2
0302			0.03175	4								
0401	0.01835	4	0.01587	2					0.02778	2		
0402	0.00459	1	0.00794	1								
0403	0.01376	3	0.01587	2							0.11111	2
0404			0.01587	2								
0405	0.00459	1	0.03175	4								
0407	0.01376	3										
0408	0.01376	3										
0411	0.01835	4	0.02381	3					0.02778	2		
0701	0.12385	27	0.07937	10	0.10000	2	0.10000	2	0.05556	4	0.05556	1
0801	0.05046	11	0.03175	4								
0802	0.01376	3	0.03175	4	0.05000	1	0.05000	1	0.01389	1		
0803			0.00794	1								
0804	0.00459	1			0.10000	2	0.10000	2	0.04167	3	0.05556	1
0807	0.00459	1			0.05000	1	0.05000	1	0.01389	1		
1001	0.01376	3	0.02381	3					0.02778	2	0.11111	2
1101	0.06422	14	0.06349	8	0.05000	1	0.05000	1	0.08333	6	0.05556	1
1102	0.01835	4			0.05000	1	0.05000	1	0.05556	4	0.05556	1
1103			0.01587	2					0.01389	1		
1104	0.03670	8	0.01587	2								

1201	0.00917	2	0.02381	3	0.05000	1	0.05000	1	0.02778	2		
1301	0.05505	12	0.03968	5					0.05556	4		
1302	0.03670	8	0.05556	7	0.05000	1	0.05000	1	0.11111	8	0.11111	2
1303	0.00917	2									0.11111	2
1401	0.05963	13	0.02381	3					0.02778	2	0.05556	1
1402	0.01835	4	0.02381	3	0.05000	1	0.05000	1	0.04167	3		
1404	0.00917	2	0.00794	1								
1406	0.01376	3									0.05556	1
1501	0.09633	21	0.03175	4	0.05000	1	0.05000	1	0.04167	3	0.05556	1
1502	0.01376	3	0.00794	1								
1503	0.02294	5	0.07937	10	0.10000	2	0.10000	2	0.11111	8	0.05556	1
1601	0.03211	7	0.00794	1								
1602	0.01376	3	0.06349	8	0.10000	2	0.10000	2	0.04167	3		
9012	0.00459	1	0.03175	4					0.01389	1		
Total		218		126		20		20		72		18

TABELA 21. Frequência HLA-DR LGMH por etnia

	Branco		Mulato Claro Europeu		Mulato Claro Brasileiro		Mulato Escuro		Mulato Medio		Negro	
HLA-DR	frel	f	frel	f	frel	f	frel	f	frel	f	frel	f
01	0.10092	22	0.09524	12	0.10000	2	0.10000	2	0.08333	6		
03	0.08257	18	0.11111	14					0.08333	6	0.11111	2
04	0.08716	19	0.11111	14					0.02778	2	0.11111	2
07	0.12385	27	0.07937	10	0.10000	2			0.05556	4	0.05556	1
08	0.07339	16	0.07143	9	0.25000	5	0.10000	2	0.08333	6	0.05556	1
09	0.00459	1	0.03175	4			0.25000	5	0.01389	1		
10	0.01376	3	0.02381	3					0.02778	2	0.11111	2
11	0.11927	26	0.10317	13	0.10000	2	0.10000	2	0.18056	13	0.11111	2
12	0.00917	2	0.02381	3	0.05000	1	0.05000	1				
13	0.10092	22	0.10317	13	0.10000	2	0.10000	2	0.18056	13	0.22222	4
14	0.10550	23	0.05556	7	0.05000	1	0.05000	1	0.06944	5	0.11111	2
15	0.13303	29	0.11905	15	0.15000	3	0.15000	3	0.15278	11	0.11111	2
16	0.04587	10	0.07143	9	0.10000	2	0.10000	2	0.04167	3		
Total		218		126		20		20		72		18

Para todas as amostras o p.valor encontrava-se em equilíbrio de HW. A análise desta tabela apresentou alguns fatores importantes, por exemplo, na análise do HLA-DRB1, para a etnia branca houve dominância do HLA-DRB1*0701 e essa tipagem obteve valores expressivos somente para esta etnia. Para a etnia mulato médio os valores de maior frequência foram HLA-DRB1*1302 e 1503. A etnia mulato claro europeu obteve destaque para a tipagem 0701, assim como na branca, mas teve um empate, obtendo também destaque na 1503 e 0301. Para a tipagem analisando somente os 2 primeiros caracteres os destaques foram para a tipagem DR*07 na etnia branca, DR*15 na etnia mulato claro com ascendência européia e mulato médio com empate em 18% para HLA-DR*11 e DR*13.

4.7 ANÁLISE DE COMPARAÇÃO ENTRE LIHC E LGMH

A análise na comparação do antígeno HLA-DRB1 para os bancos LGMH(dados painel) e para os indivíduos doadores familiares de pacientes sem anemia falciforme e sem anemia aplástica severa e também não consanguíneos foi realizada através do teste qui-quadrado. A comparação foi realizada através da frequência absoluta de cada especificidade em separado por etnias. As etnias comparadas foram Branca do LGMH com Branca do LIHC e etnia Mulato Médio do LGMH com etnia Mulata do LIHC. As demais etnias não foram consideradas nesse estudo em virtude da baixa quantidade de indivíduos.

Há análise foi realizado envolvendo as técnicas de Pearson e de Yates. Quando a frequência era inferior a 5, procurou-se utilizar a técnica de Yates, quando não a de Pearson era a utilizada. Foram constatadas diferenças nas especificidades HLA-DRB1*0402, 0404 e 1401 para a etnia Branca, enquanto para a etnia Mulata não houve nenhuma análise onde o p.valor fosse inferior a 0,05. Se considerarmos que das 43 especificidades distintas para a etnia branca, somente 3 apresentaram p.valor significativo, podemos considerar a amostra como possível de ser utilizada como controle. Porém uma ressalva deve ser feita, para a tipagem HLA-DRB1, o banco do LIHC apresentou-se fora do equilíbrio de HardyWeinberg para o algoritmo de Guo e Thompson, porém na análise de heterozigocidade esta amostra apresenta-se dentro do equilíbrio este talvez tenha sido o fator de impacto na amostragem. O

fato de não existir um grupo controle saudável com as mesmas tipagens realizadas, somente a análise de DRB1, é insuficiente para afirmar com total garantia que a amostra do laboratório de Imunogenética do Hospital das Clínicas possa ser usada sem qualquer receio. As tabelas do APÊNDICE 12 e 13 demonstram todos os dados obtidos para esta análise.

4.8 ANÁLISE PACIENTES DE PÊNFIGO, LEUCEMIA E ANEMIA DE FANCONI

Uma análise comparativa entre os pacientes de pênfigo, leucemias e anemia de Fanconi foi realizada. Os resultados estão demonstrados na sequência.

4.8.1 Pênfigo

A análise do antígeno HLA-A demonstrou que o de maior frequência foi o HLA-A*02 com frequência de 0,204, seguido pelo HLA-A*68, com 0,112 e HLA-A*03 com 0,204. Resultados estes um pouco diferentes dos dados de DVMO. Para os antígenos HLA-B, os resultados mais expressivos foram HLA-B*14;15 e 35 com frequências 0.13483; 0.12921 e 0.11798 respectivamente. Os resultados podem ser acompanhados no APÊNDICE 8 e 9.

4.8.2 Leucemia e Anemia de Fanconi

Para os antígenos HLA-A e HLA-B, observou-se uma diferenciação entre as frequências obtidas. A tabela abaixo ilustra as diferenças encontradas.

TABELA 22. Frequência HLA-A e HLA-B Leucemia e Fanconi comparativa

Leucemia			Fanconi		
Antígeno	frel	f	Antígeno	frel	f
HLA-A*0201	0.21569	44	HLA-A*0201	0.15000	15
HLA-A*2402	0.12745	26	HLA-A*2402	0.11000	11
HLA-A*0101	0.10294	21	HLA-A*0301	0.10000	10
HLA-B*0801	0.08661	22	HLA-B*4403	0.09615	10
HLA-B*1402	0.05512	14	HLA-B*3501	0.07692	8
HLA-B*3501	0.05118	13	HLA-B*1402	0.05769	6

É observável que os dois primeiros antígenos HLA-A mais comuns para ambas as doenças são iguais, porém as frequências são distintas entre si. Já para o caso do antígeno HLA-B os mais frequentes diferem totalmente tanto em especificidade quanto em frequência. As frequências para HLA-A e HLA-B com todos os valores obtidos encontram-se no APÊNDICE 10 e 11.

Para o antígeno DRB1 o resultado obtido foi o descrito na **TABELA 23**. Observa-se uma grande diferença entre as tipagens. É possível inferir que existe certa associação, para a amostragem, entre as doenças e o gene HLA-DRB1.

TABELA 23. Frequência DRB1 Pênfigo Leucemia e Anemia de Fanconi

Antígeno	Pênfigo		Leucemia		Anemia Fanconi	
	frel	f	frel	f	frel	f
HLA-DRB1*0001	0.04358	19	0.00034	1		
HLA-DRB1*0002	0.00459	2				
HLA-DRB1*0004	0.06422	28				
HLA-DRB1*0007	0.00229	1				
HLA-DRB1*0008	0.00688	3				
HLA-DRB1*0012	0.00229	1				
HLA-DRB1*0013	0.00917	4				
HLA-DRB1*0014	0.00688	3				
HLA-DRB1*0015	0.00688	3				
HLA-DRB1*0016	0.02294	10				
HLA-DRB1*0052	0.00688	3				
HLA-DRB1*0101	0.08257	36	0.05988	177	0.03650	20
HLA-DRB1*0102	0.15826	69	0.03146	93	0.03467	19
HLA-DRB1*0103	0.01376	6	0.00677	20	0.01460	8
HLA-DRB1*0104			0.00068	2	0.00182	1
HLA-DRB1*0301	0.02064	9	0.08491	251	0.09489	52
HLA-DRB1*0302	0.02982	13	0.00880	26	0.02007	11
HLA-DRB1*0304			0.00034	1		
HLA-DRB1*0313			0.00034	1		
HLA-DRB1*0315					0.00182	1
HLA-DRB1*0401	0.01606	7	0.02706	80	0.01642	9
HLA-DRB1*0402	0.02064	9	0.01894	56	0.02555	14
HLA-DRB1*0403	0.00688	3	0.01049	31	0.00912	5
HLA-DRB1*0404	0.06193	27	0.03112	92	0.02555	14
HLA-DRB1*0405	0.01147	5	0.01658	49	0.01825	10

HLA-DRB1*0406	0.01376	6	0.00304	9		
HLA-DRB1*0407			0.00846	25	0.01642	9
HLA-DRB1*0408			0.00203	6	0.00365	2
HLA-DRB1*0409	0.00229	1	0.00034	1		
HLA-DRB1*0410	0.00459	2			0.00182	1
HLA-DRB1*0411	0.04587	20	0.01793	53	0.02372	13
HLA-DRB1*0701	0.00917	4	0.12314	364	0.14781	81
HLA-DRB1*0801			0.02436	72	0.01642	9
HLA-DRB1*0802	0.01147	5	0.01556	46	0.00365	2
HLA-DRB1*0803			0.00609	18	0.00912	5
HLA-DRB1*0804	0.00459	2	0.01353	40	0.01642	9
HLA-DRB1*0806			0.00068	2		
HLA-DRB1*0807	0.00688	3	0.00744	22	0.00365	2
HLA-DRB1*0814			0.00034	1		
HLA-DRB1*0901	0.00229	1	0.01150	34	0.02372	13
HLA-DRB1*0912	0.00229	1				
HLA-DRB1*1001	0.01147	5	0.02300	68	0.01642	9
HLA-DRB1*1101	0.00229	1	0.06258	185	0.04562	25
HLA-DRB1*1102	0.00459	2	0.01455	43	0.01642	9
HLA-DRB1*1103			0.00880	26	0.00365	2
HLA-DRB1*1104	0.00229	1	0.03180	94	0.03285	18
HLA-DRB1*1119			0.00034	1		
HLA-DRB1*1201	0.01606	7	0.01455	43	0.02372	13
HLA-DRB1*1202	0.00229	1	0.00135	4		
HLA-DRB1*1203	0.00229	1	0.00034	1		
HLA-DRB1*1301	0.01376	6	0.06834	202	0.05109	28
HLA-DRB1*1302	0.04587	20	0.04432	131	0.03832	21
HLA-DRB1*1303			0.01184	35	0.01277	7
HLA-DRB1*1304			0.00068	2		
HLA-DRB1*1305			0.00101	3	0.00182	1

HLA-DRB1*1323					0.00182	1
HLA-DRB1*1356			0.00034	1		
HLA-DRB1*1401	0.01606	7	0.02740	81	0.02007	11
HLA-DRB1*1402	0.00459	2	0.01590	47	0.01460	8
HLA-DRB1*1404	0.00459	2	0.00135	4		
HLA-DRB1*1406	0.01147	5	0.00237	7	0.00365	2
HLA-DRB1*1413			0.00034	1		
HLA-DRB1*1501	0.03670	16	0.07578	224	0.07482	41
HLA-DRB1*1502			0.00981	29	0.00547	3
HLA-DRB1*1503	0.02064	9	0.01421	42	0.03650	20
HLA-DRB1*1601	0.03211	14	0.01827	54	0.00912	5
HLA-DRB1*1602	0.05963	26	0.01793	53	0.02372	13
HLA-DRB1*1603					0.00182	1
Total	100.000	436	100.000	2956	100.000	548

4.9 ANÁLISE DE GEOPOSIÇÃOAMENTO

A análise da distribuição geográfica dos indivíduos não demonstrou nenhum padrão reconhecível. Talvez a mudança na forma de plotagem dos dados seja necessária para que qualquer padrão possa ser observado, por exemplo, utilização de agrupamento por região, bairro, município, etc. Ou ainda, talvez, a criação de um algoritmo para identificação automática. Mesmo que não revelando como uma ferramenta para reconhecimento de características na população, essa ainda se revela interessante na análise e distribuição dos indivíduos. Atende a um objetivo importante que é o fato da conversão em coordenadas, agrupamentos futuros ou métodos de plotagem distintos podem ser facilmente empregados a partir deste momento.

5 DISCUSSÃO

A padronização dos dados foi imprescindível para a realização deste trabalho. A remoção de dados duplicados ou com preenchimento errôneo melhorou em muito a pesquisa. Tal fator ocorre em vista da remoção de dados considerados inválidos ou inconsistentes, o que poderia distorcer a realidade dos resultados obtidos neste trabalho. A padronização também permitiu que fosse mais fácil a comparação entre as amostras e a filtragem dos dados ocorresse mais naturalmente. Trabalhos futuros poderão ser desenvolvidos a partir deste banco, e certamente, outras conclusões serão observadas. A criação de um padrão para integração destes dados também permitirá, possivelmente a troca, de dados entre os setores LGMH e LIHC, permitindo uma maior integração das entidades. O desenvolvimento de um sistema de gerenciamento deste dados está sendo desenvolvido ao LGMH, este sistema irá alimentar ainda mais a quantidade de dados deste banco criado neste trabalho.

5.1 FREQUENCIA ALÉLICA

Em comparação de frequências entre o estudo realizado por (DALALIO *et al*, 2002) e (RUIZ *et al*, 2005) referente ao banco de dados de DVMO, foi observada praticamente a mesma frequência. Isso demonstra que não houve desequilíbrio na população de estudo, fatores como seleção natural entre outros não se mostraram significativos. A análise de Dalalio, envolveu 732 indivíduos, enquanto a de Ruiz 3500. As frequências encontradas foram para o primeiro HLA-A*02 com 0,29; HLA-A*03 com 0,12 e HLA-A*01 com 0,10 enquanto para o segundo as frequências obtidas respectivamente para os mesmos antígenos foram: 0,23 ; 0,095 e 0,09. Este trabalho obteve 0,27 ; 0,09 e 0,1. Estes dados bem próximos também demonstram que a pesquisa em ambos os trabalhos foram realizadas de maneira similar. Dados semelhantes foram obtidas também na análise do antígeno HLA-B, onde Dalalio encontrou respectivamente para HLA-B*35 , 44, e 51 os valores 0,119; 0,118 e 0,09, enquanto Ruiz obteve 0,11; 0,10 e 0,08 e este trabalho 0,11; 0,10 e 0,09. É possível visualizar na **FIGURA 23**.

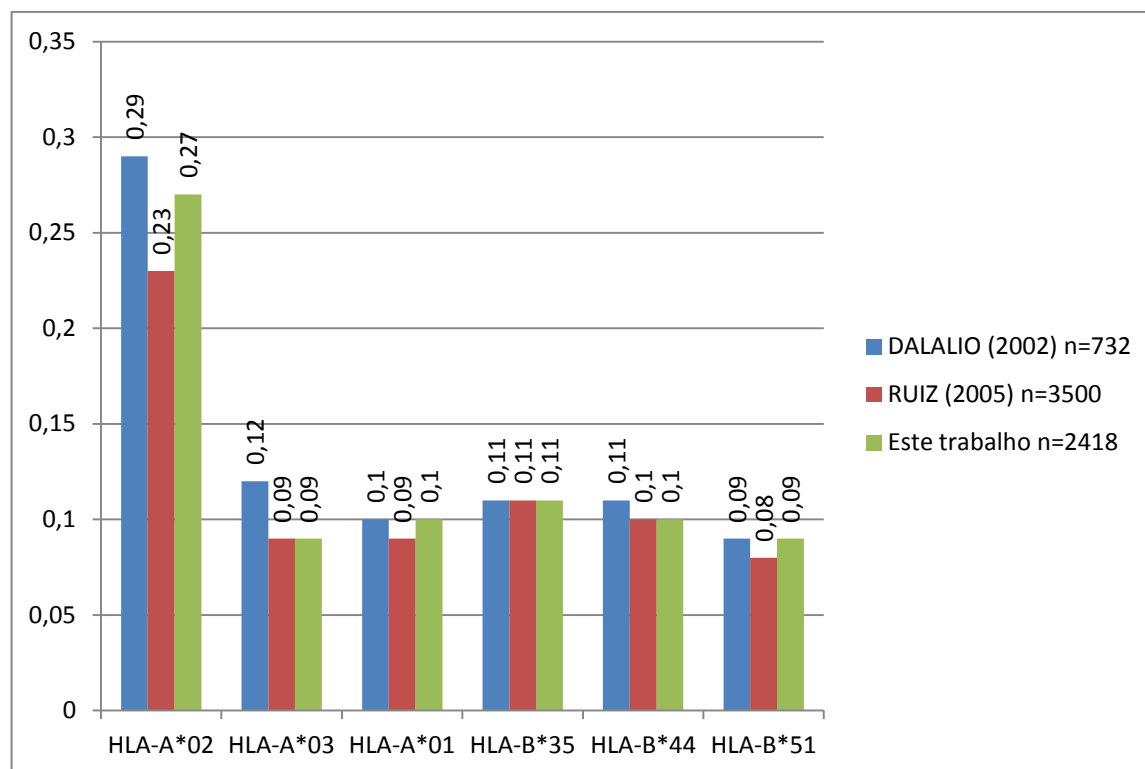



FIGURA 23 - Comparativo de trabalhos

Uma análise comparativa com o trabalho realizado por (MONTE *et al.*, 2004) e por (GOLDBERG *et al.*, 2008) também é possível. Monte realizou a pesquisa com indivíduos mestiços da região do Piauí para os antígenos HLA-A,B e DR e Goldberg com a população de Rondônia e de São Paulo, no caso a ser apresentado aqui para o antígeno HLA-DR. A **TABELA 24** ilustra os principais valores obtidos para cada um dos genes HLA em comparação com este trabalho. É possível observar que a variação entre as frequências para cada antígeno é extremamente baixa.

TABELA 24. Comparativo Antígenos Goldberg, Monte e DVMO

Antígeno	Goldbeg		Monte	DVMO	\bar{x}	
	São Paulo	Porto Velho				
HLA-A*02			0,227	0,267	0,247	0,028284271
HLA-A*03			0,121	0,093	0,107	0,01979899
HLA-A*24			0,098	0,105	0,1015	0,004949747
HLA-B*07			0,109	0,064	0,0865	0,031819805
HLA-B*15			0,098	0,091	0,0945	0,004949747
HLA-B*44			0,098	0,102	0,1	0,002828427
HLA-DR*11	0,125	0,071	0,16	0,124	0,1183	0,044769781
HLA-DR*04	0,115	0,141	0,145	0,12	0,1353	0,013428825
HLA-DR*13	0,147	0,098	0,139	0,137	0,1247	0,023115651

A análise revelou ainda que algumas tipagens para certas etnias descritas nos resultados apresentaram-se fora do equilíbrio de HW. Estes resultados ora apareceram somente no algoritmo de Guo e Thompson, ora somente para homozigocidade e afins. As frequências como um todo, pelo menos, na análise das de maior expressividade foram comuns tanto para o banco DVMO como para a amostra 3.

A análise foi bem semelhante em comparação da amostra 2 com a amostra 3. Por exemplo, as frequências calculadas para HLA-A*0201 , HLA-A*0101 e HLA-A*2402, foram respectivamente 0,21 ; 0,09 e 0,08 enquanto para o banco da amostra 3, foram 0,22 ; 0,09 e 0,08. O mesmo se deu para o antígeno B, onde para HLA-B*0702, 4403 e 0801 as frequências obtidas foram 0,07; 0,06 e 0,06 enquanto para a amostra 3 foram 0,06; 0,06 e 0,07. Houve uma inversão do mais frequente, porém as frequências estavam muito próximas pra calcular como algo totalmente diferente. A análise do antígeno DRB1 também resultou no mesmo padrão, ou seja, frequências bem próximas para os antígenos estudados, dessa forma pode-se inferir que a seleção aleatória de apenas um indivíduo de cada família não implicou em mudança significativa dos valores em comparação com a amostra 3.

Como esperado foram observados diferenças nas frequências dos antígenos entre os grupos étnicos. A análise, por exemplo, da amostra de DVMO revelou que para o antígeno A , na etnia branca a dominância da especificidade HLA-A*02 e depois HLA-A*11, enquanto para a parda ocorre dominância da especificidade 02 e depois 24. O mesmo ocorre para a análise do antígeno HLA-B, onde para pardo o terceiro mais comum é a especificidade HLA-A*15 e para a etnia branca é a 51. Outros antígenos foram encontrados somente em algumas etnias. Estes dados nos informam que a origem dos povos e o cruzamento entre si provocam a mistura dos antígenos, mas que ainda conservam em partes a sua ancestralidade.

5.2 FREQUENCIA HAPLOTÍPICA

A população brasileira é composta por uma forte miscigenação de vários grupos étnicos. Os estudos realizados e os dados obtidos mostraram uma forte associação entre o grupo étnico e as características de cada especificidade dos

antígenos. Mostrando que houve maior frequência de especificidade alélica para cada grupo étnico.

A comparação dos haplótipos mostrou dados semelhantes aos obtidos na frequência alélica, onde tivemos independente da amostra a expressão de valores haplótipos quase iguais em frequência. Para a combinação haplotípica HLA-A,B,DRB1, obteve-se para a especificidade 0101:0801:0301 a frequência de 0,045 enquanto para a especificidade 2902:4403:0701 a frequência de 0,041 para a amostra 3 esses valores foram respectivamente 0,05 e 0,049.

As frequências haplotípicas dos genes HLA-A, B e DRB1 também mostram-se divergentes entre as populações de pacientes de leucemia e de anemia de Fanconi. Dessa forma acredita-se que seja possível uma associação entre a combinação haplotípica e as doenças citadas. A **TABELA 25** demonstra estes resultados.

TABELA 25. Frequência haplotípica HLA-A;B;DRB1 Leucemia e Anemia Fanconi

Leucemia		Anemia Fanconi	
HLA-A;HLA-B;HLA-DRB1	frel	HLA-A;HLA-B;HLA-DRB1	frel
0101:0801:0301:	0.05172	2902:4403:0701:	0.03409
2402:1402:0102:	0.02874	0201:3701:0701:	0.02273
2902:4403:0701:	0.02299	2301:4403:0701:	0.02273
2402:0702:1501:	0.01724	3001:4201:0302:	0.02273
2501:0801:0301:	0.01149	3002:1402:1503:	0.02273
0201:5101:1402:	0.01149	2402:4403:0701:	0.02273
0201:5101:1104:	0.01149	0211:4802:0901:	0.02273
0201:3501:1101:	0.01149	0201:5101:0701:	0.02273
0302:0801:0301:	0.01149	3301:1402:1101:	0.01136
1101:3501:1503:	0.01149	3101:3505:1402:	0.01136

5.3 PÁGINAS DO SISTEMA

As páginas revelaram-se essenciais na criação dos arquivos utilizados no programa PyPop bem como para o programa Arlequin. Análises mais personalizadas podem ser realizadas facilmente, bastante utilizar comandos SQL

diretamente no banco ou ainda utilizar a interface desenvolvida. Filtragens iniciais podem ser realizadas e exportadas para arquivos como de planilha eletrônica ".xls" ou ".arp", arquivo do Arlequin, o que permite ao pesquisador ter mais tempo para realização de análises e pesquisa do que gastar grande parte do tempo em uma filtragem manual dos dados.

5.4 BANCO DE CONTROLE

A comparação com o banco do LGMH em confronto direto com o banco de dados de familiares do HC mostrou que a análise mesmo primeiramente apresentando-se fora de equilíbrio de HW para a etnia branca no antígeno HLA-DRB1, mesmo assim demonstrou baixo grau de diferença estatística entre as diversas especificidades encontradas. A comparação somente das etnias branca e mulato e do gene HLA-DRB1 é pouca para afirmarmos que nenhum cuidado deve ser tomado na utilização do banco do LIHC como banco de controle. Comparações para os antígenos HLA-A e HLA-B bem como os demais encontrados são necessários, afinal a quantidade de registros não pode ser facilmente descartada como banco de controle.

6 CONCLUSÃO

- Foram obtidos valores de frequências semelhantes para as tipagens HLA-A e HLA-B para a mesma população de estudo de DVMO
- Foram obtidas as frequências dos antígenos HLA do banco de dados de familiares de pacientes do HC. Também mensuradas e classificadas as amostras quanto à etnia e equilíbrio HW. Estudo este inédito para os dados do LIHC.
- Comprovado que para a tipagem HLA-DRB1 o banco de dados do LIHC encontra-se fora do equilíbrio de HW e em comparação com o banco do LGMH podendo ser usado como banco de controle, porém com ressalvas pelo tipo de análise realizada.
- O confronto entre HLA-A e HLA-B dos bancos de Pênfigo e DVMO demonstraram apresentar valores significativamente diferentes. Dessa forma é possível inferir uma associação possível entre a doença e a tipagem HLA.
- Os scripts criados permitem uma melhor classificação e filtragem dos dados, bem como um menor tempo na criação de arquivos de dados para uso em outros programas estatísticos.
- A criação de um banco unificado permite a maior integração entre o LIHC e o LGMH, bem como comparações podem ser realizadas em um curto período de tempo.
- A padronização dos dados permitem comparações, eliminação de erros e correta organização dos dados.
- Comparações de geoposicionamento, ainda que não foram observadas, com a criação do banco e a normalização, poderão ser observadas no futuro. Inclusive com comparações de frequência alélica entre os laboratórios.

BLIBLIOGRAFIA

ALVAREZ, I.; CASTRO, A. L. **HLA-B27 and immunogenetics of spondyloarthropaties**. Curr Opinion Rheumatol, v. 12, p.248-253, 2000.

ALVES, C. **Associação do sistema de histocompatibilidade humano com doenças oftalmológicas**. Arq. Bras. Oftalmol, v. 69, n. 2, p. 273-278, Abril 2006.

BALL, E.; KHAN, M. A. **HLA-B27 polymorphism**. Bone Joint Spine, v. 68, n. 5, p. 378-382, 2001.

BARQUERA, R.; ZÚÑIGA, J.; HERNÁNDEZ-DÍAZ, R.; ACUÑA-ALONZO, V.; MONTOYA-GAMA, K.; MOSCOSO, J.; TORRES-GARCÍA, D.; GARCÍA-SALAS, C.; SILVA, B.; CRUZ-ROBLES, D.; ARNAIZ-VILLENA, A.; VARGAS-ALARCÓN, G.; GRANADOS, J. **HLA class I and class II haplotypes in admixed families from several regions of Mexico**. Mol Immunol, v. 45, n. 4, p. 1171-1178, Fevereiro 2008.

BICALHO, M. G.; RUIZ, T. M.; COSTA, S. M. C.; ZACARIAS, F. R. **Haplótipos HLA mais frequentes em doadores voluntários de medula óssea de Curitiba, Paraná**. Rev. Bras. Hematol. Hemoter, v. 24, n. 4, p. 306-309, 2002.

BLANCO-GELAZ, M. A.; LOPEZ-VAZQUEZ, A.; GARCIA-FERNANDEZ, S.; MARTINEZ-BORRA, J.; GONZALEZ, S.; LOPEZ-LARREA, C. **Genetic variability, molecular evolution, and geographic diversity of HLA-B27**. Hum Immunol, v. 69, n. 9, p. 1042-1050, 2001.

BOLDT, A. B. W .; CULPI, I .; TSUNETO, L. T .;SOUZA, I. R .; KUN, J. F.; PETZL-ERLER, M. L. **Analysis of the CCR5 gene coding region diversity in five South American populations reveals two new non-synonymous alleles in Amerindians and high CCR5*D32 frequency in Euro-Brazilians**. Genetics and Molecular Biology, 2009.

BRAUN-PRADO, K.; VIEIRA MION A. L.; FARAH, N. P.; CULPI, L.; PETZLER, M. L. **HLA class I polymorphism, as characterised by PCR-SSOP, in a Brazilian exogamic population.** Tissue Antigens, v. 56, p. 417-427, 2000.

BRINEY, A. A History and Overview of Medical Geography. Medical Geography. 200-?. Disponível em: <http://geography.about.com/od/culturalgeography/a/medicalgeograph.htm> Acesso em: 01/06/2011.

CHEN, I. H.; YANG, K. L.; LEE, A.; HUANG, H. H.; LIN, P. Y.; LEE, T. D. **Low frequency of HLA-B2706 in Taiwanese patients with ankylosing spondylitis.** European Journal of Immunogenetics, v. 29, n. 5, p. 435-438, 2002.

DALAI, M. M. O.; SELL, A. M.; TODA, L. Y.; CANO, M. F. F.; SOSSAI, C. R.; FRACALOSI L. **Frequência dos antígenos HLA-A e HLA-B em populações das regiões de Curitiba e norte/nordeste do Estado do Paraná.** Acta Sci, v. 24, p. 743-751, 2002.

DE PABLO, R.; BERAÚN, Y.; NIETO, A.; CALZADA, J. E.; REMENTERÍA, M. C.; SANZ, L.; LÓPEZ-NEVOT, M. A.; MARTÍN, J. **HLA class I and class II allele distribution in the Peruvian population.** Tissue Antigens, v. 56, n. 6, p. 507-528, Dez 2000.

DONADI, E. A. **Como entender a nomenclatura e os mecanismos de associação entre os antígenos e os alelos de histocompatibilidade com as doenças.** Medicina (Ribeirão Preto), v. 33, p. 7-18, 2000.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. **Arlequin (version 3.0): An integrated software package for population genetics data analysis.** vol Bioinform Online, v. 1, p. 47-50, 2005.

FAYZ, A. A. J.; SARIA, F. W. G.; ABDULJABBAR, A. A.; MONA, R. A.; NOHA, I. H.; PIERRE, N.; KHADIJA, A. O.; AYESHA, A. M.; WASSIM, Y. A. **Contribution of Selective HLA-DRB1/DQB1 Alleles and Haplotypes to the**

Genetic Susceptibility of Type 1 Diabetes among Lebanese and Bahraini Arabs.

The Journal of Clinical Endocrinology & Metabolism, v. 90, n. 9, p. 5104-5109, 2005.

FERREIRA, L. B.; MENDES-JUNIOR, C. T.; WIEZEL, C. E. V.; LUIZON, M. R.; SIMÕES A. L. **Y-STR diversity and ethnic admixture in White and Mulatto Brazilian population samples.** Genet. Mol. Biol, v. 29, n. 4, p. 605-607, 2006.

GOLDBERG, A. C .; CHIARELLA, J. M.; MARIN, M. L. C.; ROSALES, C.; BANIC, D.; OLIVEIRA, M. A.; RODRIGUES H.; VIGGIANI, C. S.; KALIL, J. **Molecular typing of HLA class II antigens in a São Paulo population.** Genet. Mol Biol, v. 21, n.3, 1998.

GUEDE,S M. E. S.; PACHECO, M. J. A.; FERNANDO, S. I. A. **Cloranfenicol.** Universidade do Porto. Portugal, 2007.

GUO, S. W.; THOMPSON, E. A. **Performing the exact test of Hardy-Weinberg proportion for multiple alleles.** Biometrics. Washington, v. 42, n.2, p. 361-372, 1992.

INFANTE, E.; OLIVO, A.; ALAEZ, C.; WILLIAMS, F.; MIDDLETON, D.; DE LA ROSA, G.; PUJOL, M. J.; DURÁN, C.; NAVARRO, J. L.; GORODEZKY, C. **Molecular analysis of HLA class I alleles inthe Mexican Seri Indians: implications for their origin.** Tissue Antigens, v. 54, n.1, p. 35-42, Jul 1999.

ITO, T.; CHIKU, S.; INOUE, E.; TOMITA, M.; MORISAKI, T.; MORISAKI, H.; KAMATANI, N. **Estimation of haplotype frequencies, linkage-disequilibrium measures, and combination of haplotype copies in each pool by use of pooled DNA data.** American Journal of Human Genetics, v. 72, n. 2, p. 384-398, Fev, 2003.

LANCASTER, A. K.; NELSON, M. P.; SINGLE, R. M.; MEYER, D.; THOMSON, G . **PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data.** Pacific Symposium on Biocomputing, v. 8, p. 514-525, 2003.

LANCASTER, A. K.; SINGLE, R. M.; SOLBERG, O. D.; NELSON, M. P.; THOMSON, G. **PyPop update - a software pipeline for large-scale multilocus population genomics**. Tissue Antigens, v. 69, n. 1, p. 192-197, 2007.

MCMANUS, D. P.; ROSS, A. G.; WILLIAMS, G. M.; SLEIGH, A. C.; WIEST, P.; ERLICH, H.; TRACHTENBERG, E.; GUANLING, W.; MCGARVEY, S. T.; LI, Y. S.; WAINE, G. J. **HLA class II antigens positively and negatively associated with hepatosplenic schistosomiasis in a Chinese population**. Int J Parasitol, v. 31, n. 7, p. 674-680, 15 Mai 2001.

MONTE, S. J. H.; MOITA NETO, J. M.; RAMPIM, G. F.; SHULZHENKO, N.; MORGUN, A. G. D. L. M. **Polimorfismo do sistema HLA em uma amostra de mestiços da população de Teresina, Piauí**. Rev. Assoc. Med. Bras, v. 50, n. 4, p. 422-426, Dez 2004.

MORAES, M. E.; FERNANDEZ-VIÑA, M.; SALATIEL, I.; TSAI, S.; MORAES, J. R.; STASTNY, P. **HLA class II DNA typing in two Brazilian populations**. Tissue Antigens, v. 41, n. 5, p. 238-242, Mai 1993.

MORI, M.; BEATTY, P. G.; GRAVES, M.; BOUCHER, K. M.; MILFORD, E. L. **HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry**. Transplantation, v. 64, n. 7, p. 1017-1027, 15 out 1997.

NAITO, S. **The association of HLA with diseases in Japanese**. Journal of Human Genetics, v. 31, n. 4, p. 323-329, Dez 1986.

OLIVEIRA, L. A. **Diversidade populacional de microssatélites do complexo principal de histocompatibilidade (mhc) em populações indígenas do paran  e do mato grosso do sul**. UFP, 2007.

PAROLIN, M. L.; CARNESE, F. R. **HLA-DRB1 alleles in four Amerindian populations from Argentina and Paraguay**. Genet. Mol. Biol, v. 32, n. 2, p. 212-219, 2009.

PARRA, E. J.; MARCINI, A.; AKEY, J.; MARTINSON, J.; BATZER, M. A.; COOPER, R.; FORRESTER, T.; ALLISON, D. B.; DEKA, R.; FERRELL, R. E.; SHRIVER, M. D. **Estimating African American admixture proportions by use of population-specific alleles**. Am J Hum Genet, v. 63, n. 6, p. 1839-1851, Dez 1998.

PERRIN, J. G. **Php and ids**. DB2 Magazine, v. 11, n. 3, p. 58-60, 2006.

PRADO, K. B. **Diversidade Molecular dos Genes HLA de Classe I em Populações Paranaenses**. Dissertação de Mestrado em Genética. Univ. Fed. do Paraná, 1999

PROBST, C. M.; BOMPEIXE, E. P.; PEREIRA, N. F.; DALALIO, M. M. O.; VISENTAINER, J. E. L.; TSUNETO, L. T.; PETZL-ERLER, M. L. **HLA Polymorphism and Evaluation of European, African and Amerindian Contribution to the White and Mulatto Populations from Paraná, Brazil**. Human Biology, v. 72, p. 597-617, 2000.

RAITZ, R. T. **FREE ASSOCIATIVE NEURONS - FAN UMA ABORDAGEM PARA RECONHECIMENTO DE PADRÕES**. Dissertação de Mestrado. UFSC, 1997.

RAMOS, J. A. S.; ESPERANCA, C.; CLUA, E. W. G. **A progressive vector map browser for the web**. J. Braz. Comp. Soc., Campinas, v. 15, n. 2, June 2009

RUIZ, T. M.; COSTA, S. M. C. M.; RIBAS, F.; LUZ, P. R.; LIMA, S. S.; BICALHO, M. G. **Human Leukocyte Antigen Allelic Groups and Haplotypes in a Brazilian Sample of Volunteer Donors for Bone Marrow Transplant in Curitiba, Paraná, Brazil**. Transplantation Proceedings, v. 37, n. 5, p. 2293-2296, Jun 2005.

SAMPAIO-BARROS, P. D.; CONDE, R. A.; DONADI, E. A.; KRAEMER, M. H. S.; PERSOLI, L.; COIMBRA, I. B.; COSTALLAT, L. T. L.; SAMARA, A. M.; BÉRTOLO, M. B. **Undifferentiated Spondyloarthropathies Brazilians: Importance of the HLA-27 and the B7-CREG alleles in the characterization and disease progression**. J Rheumatol, v. 30, n. 2, p. 2632-2637, 2003.

SANCHEZ-MAZAS, A. **African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection.** *Human immunology*, v. 62, n. 9, p. 937-948, 2001.

SANCHEZ-VELASCO, P.; KARADSHEH, N. S.; GARCIA-MARTIN, A.; ALEGRIA, C. R.; LEYVA-COBIAN, F. **Molecular analysis of HLA allelic frequencies and haplotypes in jordanians and comparison with other related populations.** *Human Immunology*, v. 62, n. 9, p. 901-909, Set 2001.

SANTOS, P. S. S.; LIMA, R. B.; MAGALHÃES, M. H. C. G. **Doença do Enxerto-Contra-Hospedeiro (DECH) em pacientes transplantados de medula óssea – relato de caso.** *RPG Rev Pós Grad, São Paulo*, v. 12 , n. 4, p.506-511, 2005.

SELL, A. M.; VISENTAINER, J. E. L.; BEZERRA, R. C.; SILVA, C. C.; BORELLI, S. C.; MOLITERNO, R. A.; TSUNETO, L. T.; DALALIO, M. M. O. **Importância da histocompatibilidade HLA no sucesso do transplante de medula óssea.** *IV Fórum de Extensão e Cultura da UEM: Perspectivas da Extensão Universitária e da Prestação de Serviços, Arq Mundi*, v. 10, n. 1, 2006.

SEVER, L. E. **Guidelines for Conducting : Birth Defects Surveillance.** National Birth Defects Prevention Network, 2004. Disponível em: http://www.nbdpn.org/docs/NBDPN_Guidelines2008.pdf . Acessado em: 10/05/2011

SMOUSE, P. E.; WILLIAMS, R. C. **Multivariate Analysis of HLA-Disease Associations.** *Biometrics. Special Issue: Analysis of Covariance*, v. 38, n. 3, p. 757-768, 1982.

SWAIN, T. N. **Fronteiras do Paraná: da colonização à migração.** *Office de la recherche scientifique et technique Outre-mer*, v. 22, n. 3, p. 269-280, 1986.

STONEBRAKER, M.; HANSON, E.; POTAMIANOS, S. **The POSTGRES rule manager.** *IEEE Trans on Software Engineering*, v. 14, n. 7, p. 897-907, Jul 1988.

STONEBRAKER, M.; JHINGRAN, J. G.; POTAMIANOS, S. **On rules, procedures, caching and views in data base systems**. Proc of ACM SIGMOD Conf on Management of Data, p. 281-290, 1975.

STONEBRAKER, M.; KEMNITZ, G. **The POSTGRES nextgeneration database management system**. Comm of the ACM, v. 34, n. 10, p. 78–92, 1991.

STONEBRAKER, M.; ROWE, A.; HIROSHIMA, M. **The implementation of POSTGRES**. IEEE Trans on Knowledge and Data Engineering, v. 2, n. 1, p. 125-142, Jul 1990.

STONEBRAKER, M.; ROWE, L. **The design of POSTGRES**. Proc of ACM-SIGMOD Conf on Management of Data, p.289-300, 1986.

STONEBRAKER, M. **The integration of rule systems and database systems**. IEEE Trans on Knowledge and Data Engineering, v. 4, n. 5, p. 415-423, 1992.

TAMOUZA, R.; MANSOUR, I.; BOUGUACHA, N.; KLAYME, S.; DJOUADI, K.; LAOUSSADI, S.; AZOURY, M.; DULPHY, N.; RAMASAWMY, R.; KRISHNAMOORTHY, R.; TOUBERT, A.; NAMAN, R.; CHARRON, D. **A new HLA-B27 allele (B*2719) identified in a Lebanese patient affected with ankylosing spondylitis**. Tissue Antigens, v. 58, n. 1, p. 30-33, 2001.

TIAGO, H. C. **Comandos básicos em SQL - insert, update, delete e select**. Linha de Código, 2010. Disponível em: <http://www.linhadecodigo.com.br/artigo/2975/Comandos-básicos-em-SQLinsert-update-delete-e-select.aspx> . Acessado em: 10/06/2011

TSUNETO, L. T.; PROBST, C. M.; HUTZ, M. H.; SALZANO, F. M.; RODRIGUEZ-DELFIN, L. A.; ZAGO, M. A.; HILL, K.; HURTADO, A. M.; RIBEIRO-DOS-SANTOS, A. K.; PETZL-ERLER, M. L. **HLA class II diversity in seven Amerindian populations. Clues about the origins of the Aché**. Tissue Antigens, v. 62, n. 6, p. 512-526, Dez 2003.

VARGAS-ALARCÓN, G.; HERNÁNDEZ-PACHECO, G.; MOSCOSO, J.; PÉREZ-HERNÁNDEZ, N.; MURGUÍA, L. E.; MORENO, A.; SERRANO-VELA, J. I.; GRANADOS, J.; ARNAIZ-VILLENA, A. **HLA genes in Mexican Teeneks: HLA genetic relationship with other worldwide populations**. Mol Immunol, v. 43, n. 7, p. 790-799, Mar 2006.

VARGAS-ALARCÓN, G.; HERNÁNDEZ-PACHECO, G.; ZUÑIGA, J.; RODRÍGUEZ-PÉREZ, J. M.; PÉREZ-HERNÁNDEZ, N.; RANGEL, C.; VILLARREAL-GARZA, C.; MARTÍNEZ-LASO, J.; GRANADOS J.; ARNAIZ-VILLENA A. **Distribution of HLA-B alleles in Mexican Amerindian populations**. Immunogenetics, v. 54, n. 11, p. 756-760, Feb 2003.

ZHANG, P.; SHENG, H.; MORABIA, A.; GILLIAM, C. T. **Optimal Step Length EM Algorithm (OSLEM) for the estimation of haplotype frequency and its application in lipoprotein lipase genotyping**. BMC Bioinformatics, v. 4, n. 3, 2003.

APÊNCICE

APÊNDICE 1 – DECLARAÇÃO ORIENTADOR

DECLARAÇÃO

Declaro estar ciente e de acordo com a apresentação do projeto de pesquisa intitulado "UM ESTUDO DE POPULAÇÕES ABORDANDO O SISTEMA DE ANTÍGENOS LEUCOCITÁRIOS HUMANOS" sob minha responsabilidade e com a autoria do aluno Luiz Fernando da Silva Soares do curso de Pós Graduação em Bioinformática sob orientação de Jeroniza Nunes Marchaukoski e co-orientação de Maria Berenice R. Steffens.

Declaro também que li e entendi a resolução 196/96 MS, responsabilizando-me pelo andamento, realização e conclusão deste projeto, comprometendo-me a enviar ao CEP/HC, relatório do presente projeto quando da sua conclusão, ou a qualquer momento, se o estudo for interrompido.

Data:14/06/10

Noemi Farah Pereira

Responsável pelo Laboratório de Imunogenética do Hospital de Clínicas

Jeroniza Nunes Marchaukoski

Coordenadora Programa de Pós Graduação em Bioinformática

APÊNDICE 2 – MODELO CARTA ENCAMINHAMENTO

Curitiba, 14 de junho de 2010

Ao Comitê de Ética em Pesquisa do HC/UFPR

Prezado Coordenador:

Encaminho a Vossa Senhoria o projeto de pesquisa intitulado “UM ESTUDO DE POPULAÇÕES ABORDANDO O SISTEMA DE ANTÍGENOS LEUCOCITÁRIOS HUMANOS” sob minha responsabilidade, para análise e parecer. Este projeto será realizado no laboratório de Bioinformática sito a Rua Dr Alcides Vieira Arcoverde, nº 1225 , no período de 16 de agosto de 2010 a 28 de fevereiro de 2011. Nossos sujeitos de pesquisa serão doadores voluntários de medula óssea e doadores familiares e não consanguíneos de pacientes do Laboratório de Imunogenética do HC

Atenciosamente,

Luiz Fernando da Silva Soares
UFPR, Setor de Educação Profissional e
Tecnológica/Programa de Pós
Graduação em Bioinformática
Pesquisador Principal

APÊNDICE 3 – MODELO DE DISPENSA DE CONSENTIMENTO

Curitiba, 14 de junho de 2010

Ao: Comitê de Ética em Pesquisa em Seres Humanos do HC / UFPR
Prof. Dr. Renato Tambara Filho

Prezado Coordenador:

Solicito a dispensa do Termo de Consentimento Livre e Esclarecido, referente ao projeto de pesquisa clínica intitulado: “UM ESTUDO DE POPULAÇÕES ABORDANDO O SISTEMA DE ANTÍGENOS LEUCOCITÁRIOS HUMANOS”, em virtude da dificuldade de localização dos 35000 indivíduos a serem estudados. A dificuldade em localização é acarretada por diversos fatores, sendo eles: Omissão de endereço, dados desatualizados, incorretos e afins. Também é justificada visto que muitos dos indivíduos a serem estudados foram a óbito e muitos dos outros a localização envolve diversas regiões do território nacional, sendo assim, e pela quantidade numérica expressiva torna-se inviável a procura destes pelos motivos já supra citados. A pesquisa irá preservar pelo completo anonimato dos indivíduos, sendo estes identificáveis por simples numeração iniciada em 1. A razão desta pesquisa tem como propósito a elaboração de dissertação como quesito parcial para obtenção do título de mestre em Bioinformática pela Universidade Federal do Paraná e permitir demais estudos de outros pesquisadores, visto que é de interesse do laboratório ter esta pesquisa realizada.

Atenciosamente,

Luiz Fernando da Silva Soares
Pesquisador Principal

APÊNDICE 4 – MODELO IDENTIFICAÇÃO DE PROJETO

IDENTIFICAÇÃO DO PROJETO

☐ PROJETO PESQUISA

☒ TESE MESTRADO

☐ TESE DOUTORADO

☐ CONCLUSÃO DE CURSO

☐ MONOGRAFIA

☐ PROSPECTIVO

☐ RETROSPECTIVO

SETOR DE ORIGEM: UNIVERSIDADE FEDERAL DO PARANÁ - UFPR

CURSO: Pós Graduação em Bioinformática

DISCIPLINA: Tese de Mestrado

DEPARTAMENTO: Setor de Educação Profissional e Tecnológica/Programa de Pós Graduação em Bioinformática

"UM ESTUDO DE POPULAÇÕES ABORDANDO O SISTEMA DE ANTÍGENOS LEUCOCITÁRIOS HUMANOS"

LOCAL A SER REALIZADO: RUA DR. ALCIDES VIEIRA ARCOVERDE, NÚMERO 1225

PERÍODO: de 16/08/2010 a 28/02/2011

Nome do pesquisador : Luiz Fernando da Silva Soares

Fones/ Celular: (41) 3035-3170 / (41) 9619-9504

e-mail: luizsoares1988@hotmail.com

ORIENTADOR: Jeroniza Nunes Marchaukoski

Co- Autores: Maria Berenice R. Steffens, Noemi F. Pereira e Maria Luiza Petzel-Erler

Local data

Curitiba, 14 de junho de 2010

APÊNDICE 5 – FOLHA DE ROSTO COMITÊ



MINISTÉRIO DA SAÚDE
Conselho Nacional de Saúde
Comissão Nacional de Ética em Pesquisa - CONEP

FOLHA DE ROSTO PARA PESQUISA ENVOLVENDO SERES HUMANOS					FR - 355847
Projeto de Pesquisa UM ESTUDO DE POPULAÇÕES ABORDANDO O SISTEMA DE ANTÍGENOS LEUCOCITÁRIOS HUMANOS					
Área de Conhecimento 2.00 - Ciências Biológicas - 2.02 - Genética				Grupo Grupo II	Nível
Área(s) Temática(s) Especial(s) Genética Humana,				Fase Não se Aplica	
Unitermos HLA, REDE NEURONAL, NEURAL, MLP, ANTÍGENOS, POPULAÇÕES, CURITIBA, PARANÁ					
Sujeitos na Pesquisa					
Nº de Sujeitos no Centro 35000	Total Brasil 35000	Nº de Sujeitos Total 35000	Grupos Especiais		
Placebo NÃO	Medicamentos HIV / AIDS NÃO	Wash-out NÃO	Sem Tratamento Específico NÃO	Banco de Materiais Biológicos NÃO	
Pesquisador Responsável					
Pesquisador Responsável LUIZ FERNANDO DA SILVA SOARES			CPF 066.760.019-18	Identidade 81673071	
Área de Especialização EXATAS			Maiores Titulação GRADUAÇÃO	Nacionalidade BRASILEIRO	
Endereço R.DR.CLAUDINO DOS SANTOS, 750 AP:504			Bairro CENTRO	Cidade SÃO JOSÉ DOS PINHAIS - PR	
Código Postal 83005-180	Telefone 41-96199604 / 41-30353170		Fax	Email luizsoares1988@hotmail.com	
<p>Termo de Compromisso</p> <p>Declaro que conheço e cumpro os requisitos da Res. CNS 196/96 e suas complementares. Comprometo-me a utilizar os materiais e dados coletados exclusivamente para os fins previstos no protocolo e publicar os resultados sejam eles favoráveis ou não.</p> <p>Aceito as responsabilidades pela condução científica do projeto acima.</p> <p>Data: ____/____/____ Assinatura _____</p>					
Instituição Onde Será Realizado					
Nome Hospital de Clínicas - Universidade Federal do Paraná - UFPR		CNPJ 75.095.679/0002-20	Nacional/Internacional Nacional		
Unidade/Órgão Laboratório de Imunogenética		Participação Estrangeira NÃO	Projeto Multicêntrico NÃO		
Endereço Rua General Carneiro, 181		Bairro Centro	Cidade Curitiba - PR		
Código Postal 80060-900	Telefone (41) 3601896		Fax (41) 3632927	Email cep@hc.ufpr.br	
<p>Termo de Compromisso</p> <p>Declaro que conheço e cumpro os requisitos da Res. CNS 196/96 e suas complementares e como esta instituição tem condições para o desenvolvimento deste projeto, autorizo sua execução.</p> <p>Nome: _____ Assinatura _____</p> <p>Data: ____/____/____</p>					

APÊNDICE 6 - HAPLÓTIPO SIMPLES

[Profile]

Title="Insira um titulo da busca"
NbSamples=1

GenotypicData=1
GameticPhase=0
DataType=STANDARD
LocusSeparator= WHITESPACE
MissingData='?'
RecessiveData=X

#Try different epsilon value (in General Settings dialog box)
#and different values of No. of initial conditions
#(in the Diversity Indices dialog box)
#for the estimation of haplotype frequencies
#It should have an impact on the final likelihood of the sample

RecessiveAllele="null"

[Data]

[[Samples]]

		SampleName="Amostra Personalizada"
		SampleSize= 195
		SampleData={
1	84	? ? ?
		? ? ?
2	3	1101 0301 0102
		1501 0602 0501
3	2	0101 02 0201
		0701 0501 0101
4	2	0101 0501 0101
		0101 0501 0101

FIGURA 24 - Haplótipo simples

APÊNDICE 7 - HAPLÓTIPO DUPLO

```

[[Samples]]
    SampleName="Amostra Personalizada"
    SampleSize= 195
    SampleData={
1      84      ?      ?      ?
2      21      0701    02    0201
3      16      0101    0501    0101
4      16      0301    02    0501
5      16      1501    0602    0102
6      8       0101    0501    (0101)
7      8       1101    0301    0501
8      7       1501    0602    0501

```

FIGURA 25 - Haplótipo duplo

APÊNDICE 8 - FREQUÊNCIA HLA-A PÊNFIGO

TABELA 26. Frequência HLA-A Pênfigo

Antígeno	frel	f
HLA-A*01	0.05914	11
HLA-A*02	0.20430	38
HLA-A*03	0.10753	20
HLA-A*11	0.07527	14
HLA-A*23	0.06452	12
HLA-A*24	0.08602	16
HLA-A*29	0.02688	5
HLA-A*30	0.03763	7
HLA-A*31	0.09140	17
HLA-A*32	0.04839	9
HLA-A*33	0.05376	10
HLA-A*34	0.00538	1
HLA-A*36	0.01075	2
HLA-A*68	0.11290	21
HLA-A*74	0.01075	2
HLA-A*80	0.00538	1
Total	100.000	186

APÊNDICE 9 - FREQUÊNCIA HLA-B PÊNFIGO

TABELA 27. Frequência HLA-B Pênfigo

Antígeno	frel	f
HLA-B*07	0.08989	16
HLA-B*08	0.01124	2
HLA-B*13	0.00562	1
HLA-B*14	0.13483	24
HLA-B*15	0.12921	23
HLA-B*18	0.02247	4
HLA-B*27	0.01124	2
HLA-B*35	0.11798	21
HLA-B*37	0.00562	1
HLA-B*39	0.07303	13
HLA-B*40	0.07303	13
HLA-B*41	0.01685	3
HLA-B*42	0.02247	4
HLA-B*44	0.08989	16
HLA-B*45	0.02809	5
HLA-B*47	0.00562	1
HLA-B*48	0.02809	5
HLA-B*49	0.00562	1
HLA-B*50	0.01685	3
HLA-B*51	0.02809	5
HLA-B*52	0.01685	3
HLA-B*53	0.01124	2
HLA-B*55	0.01124	2
HLA-B*56	0.00562	1
HLA-B*57	0.02247	4
HLA-B*58	0.00562	1
HLA-B*73	0.00562	1
HLA-B*82	0.00562	1
	100.000	178

APÊNDICE 10 - FREQUÊNCIA HLA-A LEUCEMIA E ANEMIA DE FANCONI

TABELA 28. Frequência HLA-A Leucemia e Anemia de Fanconi

Antígeno	Leucemia		Anemia Fanconi	
	frel	f	frel	f
HLA-A*0101	0.10294	21	0.09000	9
HLA-A*0201	0.21569	44	0.15000	15
HLA-A*0202	0.00980	2	0.03000	3
HLA-A*0204			0.01000	1
HLA-A*0205	0.02451	5	0.01000	1
HLA-A*0206	0.00490	1		
HLA-A*0211	0.00980	2	0.02000	2
HLA-A*0212	0.00490	1		
HLA-A*0217	0.00490	1		
HLA-A*0220	0.00490	1		
HLA-A*0301	0.04412	9	0.10000	10
HLA-A*0302	0.01471	3		
HLA-A*1101	0.04412	9	0.07000	7
HLA-A*2301	0.02451	5	0.07000	7
HLA-A*2402	0.12745	26	0.11000	11
HLA-A*2403	0.00490	1	0.01000	1
HLA-A*2501	0.02451	5	0.01000	1
HLA-A*2601	0.01961	4	0.01000	1
HLA-A*2608	0.00490	1	0.01000	1
HLA-A*2901	0.01471	3		
HLA-A*2902	0.06863	14	0.06000	6
HLA-A*3001	0.01471	3	0.06000	6
HLA-A*3002	0.00490	1	0.04000	4
HLA-A*3004	0.00490	1	0.05000	5
HLA-A*3101	0.06863	14		
HLA-A*3201	0.03431	7		
HLA-A*3301	0.01471	3	0.03000	3
HLA-A*3303	0.00490	1		
HLA-A*3402	0.00490	1		
HLA-A*3601	0.00490	1		
HLA-A*6601			0.01000	1
HLA-A*6801	0.04412	9	0.03000	3
HLA-A*6802	0.00980	2	0.01000	1
HLA-A*6901	0.00490	1		
HLA-A*7401	0.00980	2	0.01000	1
Total	100.000	204	100.000	100

APÊNDICE 11 - FREQUÊNCIA HLA-B LEUCEMIA E ANEMIA DE FANCONI

TABELA 29. Frequência HLA-B Leucemia e Anemia de Fanconi

Antígeno	Leucemia		Anemia Fanconi	
	frel	f	frel	f
HLA-B*0702	0.04724	12	0.03846	4
HLA-B*0705	0.01181	3		
HLA-B*0801	0.08661	22	0.02885	3
HLA-B*1302	0.01575	4	0.00962	1
HLA-B*1401	0.03543	9		
HLA-B*1402	0.05512	14	0.05769	6
HLA-B*1403			0.00962	1
HLA-B*1501	0.03150	8	0.00962	1
HLA-B*1503	0.00787	2	0.01923	2
HLA-B*1504	0.00394	1	0.00962	1
HLA-B*1508	0.01181	3		
HLA-B*1510	0.01575	4	0.00962	1
HLA-B*1516	0.00394	1		
HLA-B*1517	0.01575	4		
HLA-B*1518	0.00394	1		
HLA-B*1520			0.00962	1
HLA-B*1801	0.02756	7	0.04808	5
HLA-B*1805			0.00962	1
HLA-B*1804	0.00394	1		
HLA-B*2702	0.01181	3		
HLA-B*2705	0.01969	5	0.00962	1
HLA-B*3501	0.05118	13	0.07692	8
HLA-B*3502	0.01575	4	0.00962	1
HLA-B*3503	0.01969	5	0.03846	4
HLA-B*3504	0.00787	2		
HLA-B*3505	0.01969	5	0.01923	2
HLA-B*3508	0.00394	1		
HLA-B*3509	0.00394	1		
HLA-B*3510	0.00394	1		
HLA-B*3520			0.00962	1
HLA-B*3701	0.01575	4	0.01923	2
HLA-B*3801	0.01181	3	0.00962	1
HLA-B*3901	0.00394	1		
HLA-B*3903	0.01575	4		
HLA-B*3906	0.00394	1		
HLA-B*4001	0.03150	8		
HLA-B*4002	0.02756	7	0.02885	3
HLA-B*4003	0.00394	1		
HLA-B*4004	0.01575	4	0.00962	1

HLA-B*4006	0.01181	3		
HLA-B*4027	0.00394	1		
HLA-B*4101	0.00787	2	0.00962	1
HLA-B*4102			0.02885	3
HLA-B*4201	0.00394	1	0.01923	2
HLA-B*4402	0.01969	5	0.01923	2
HLA-B*4403	0.03543	9	0.09615	10
HLA-B*4404			0.00962	1
HLA-B*4501	0.01575	4		
HLA-B*4601	0.00394	1		
HLA-B*4701	0.00394	1	0.00962	1
HLA-B*4801	0.00394	1		
HLA-B*4802	0.00787	2	0.02885	3
HLA-B*4901	0.03543	9	0.04808	5
HLA-B*5001	0.02362	6	0.03846	4
HLA-B*5002	0.00787	2		
HLA-B*5101	0.04724	12	0.02885	3
HLA-B*5108			0.00962	1
HLA-B*5201	0.02362	6	0.02885	3
HLA-B*5202	0.00394	1		
HLA-B*5301	0.01575	4	0.02885	3
HLA-B*5501	0.00787	2	0.01923	2
HLA-B*5502	0.00394	1		
HLA-B*5507	0.00394	1		
HLA-B*5601	0.00394	1	0.00962	1
HLA-B*5701	0.01575	4	0.01923	2
HLA-B*5702			0.00962	1
HLA-B*5703	0.00787	2	0.00962	1
HLA-B*5801	0.01181	3	0.03846	4
HLA-B*8101	0.00394	1		
Total	100.000	254	100.000	104

APÊNDICE 12 - TABELA COMPARATIVA DRB1 BRANCO LIHC e LGMH

TABELA 30. Tabela comparativa DRB1 etnia branca - LIHC e LGMH

Branco		Yates			Pearson		
HLA-DRB1	Análise Recomendada	Qui quadrado	Valor de P		Qui Quadrado	Valor de P	
0101	Pearson	0,581071363	0,445892692	Não há diferença	0,859637361	0,353840198	Não há diferença
0102	Yates	3,422390473	0,06431783	Não há diferença	4,513540122	0,033627577	Há diferença *
0103	Yates	0,790533999	0,373938707	Não há diferença	2,178129043	0,139984589	Não há diferença
0104	Yates	0,026845477	0,869852461	Não há diferença	0,723245161	0,395081497	Não há diferença
0301	Pearson	0,139228959	0,709048738	Não há diferença	0,278899896	0,597423173	Não há diferença
0302	Yates	0,026845477	0,869852461	Não há diferença	0,723245161	0,395081497	Não há diferença
0401	Yates	0,011137998	0,915949938	Não há diferença	0,217700043	0,640797987	Não há diferença
0402	Yates	5,75543308	0,016437744	Há diferença	7,148069348	0,007504473	Há diferença *
0403	Yates	0,112341801	0,73749362	Não há diferença	0,002542625	0,959784153	Não há diferença
0404	Yates	5,709159659	0,016876617	Há diferença	7,360083106	0,006668771	Há diferença *
0405	Yates	0,032389621	0,857175203	Não há diferença	0,474149767	0,491084252	Não há diferença
0406	Yates	0,026845477	0,869852461	Não há diferença	0,723245161	0,395081497	Não há diferença
0407	Yates	0,135269929	0,713029375	Não há diferença	0,677576585	0,41042247	Não há diferença
0408	Yates	0,700999571	0,402448029	Não há diferença	1,811369666	0,1783441	Não há diferença
0409	Yates	0,026845477	0,869852461	Não há diferença	0,723245161	0,395081497	Não há diferença
0411	Yates	0,039646942	0,842172476	Não há diferença	0,015487087	0,900961264	Não há diferença
0701	Pearson	0,06553774	0,797948213	Não há diferença	0,155312066	0,693509781	Não há diferença
0801	Pearson	0,955708123	0,328270597	Não há diferença	1,460890698	0,226788568	Não há diferença
0802	Yates	2,1251674	0,144897004	Não há diferença	4,180079144	0,040901795	Há diferença *
0804	Yates	0,032389621	0,857175203	Não há diferença	0,474149767	0,491084252	Não há diferença
0807	Yates	0,032389621	0,857175203	Não há diferença	0,474149767	0,491084252	Não há diferença
0901	Yates	0,236159536	0,626993765	Não há diferença	1,449282774	0,228643194	Não há diferença

1001	Yates	1,692321104	0,193295381	Não há diferença	2,481984133	0,115156938	Não há diferença
1101	Pearson	0,014879745	0,902912828	Não há diferença	0,003634912	0,951924482	Não há diferença
1102	Yates	0,190116216	0,662819909	Não há diferença	0,675062575	0,411292112	Não há diferença
1103	Yates	1,433286659	0,231228935	Não há diferença	2,909800298	0,088042793	Não há diferença
1104	Pearson	0,000503485	0,9820982	Não há diferença	0,048684161	0,825368956	Não há diferença
1201	Yates	1,199319998	0,273457848	Não há diferença	2,012701794	0,155987364	Não há diferença
1203	Yates	0,026845477	0,869852461	Não há diferença	0,723245161	0,395081497	Não há diferença
1301	Pearson	0,397570725	0,52834651	Não há diferença	0,656568351	0,417774066	Não há diferença
1302	Yates	2,136116803	0,143865705	Não há diferença	3,088798237	0,078833028	Não há diferença
1303	Yates	0,00016391	0,989785174	Não há diferença	0,183948024	0,668002098	Não há diferença
1331	Yates	0,026845477	0,869852461	Não há diferença	0,723245161	0,395081497	Não há diferença
1401	Pearson	4,494810492	0,033997887	Há diferença *	5,502889044	0,018985084	Há diferença
1402	Yates	0,190116216	0,662819909	Não há diferença	0,675062575	0,411292112	Não há diferença
1404	Yates	0,90218798	0,342195716	Não há diferença	2,781339662	0,095368391	Não há diferença
1406	Yates	0,700999571	0,402448029	Não há diferença	1,811369666	0,1783441	Não há diferença
1501	Pearson	0,266975639	0,605367793	Não há diferença	0,455281225	0,499837597	Não há diferença
1502	Yates	0,700999571	0,402448029	Não há diferença	1,811369666	0,1783441	Não há diferença
1503	Yates	0,245391829	0,620338689	Não há diferença	0,699065477	0,403097887	Não há diferença
1601	Pearson	0,012616609	0,910566767	Não há diferença	0,142744939	0,705567278	Não há diferença
1602	Yates	0,112341801	0,73749362	Não há diferença	0,002542625	0,959784153	Não há diferença
9012	Yates	0,026845477	0,869852461	Não há diferença	0,723245161	0,395081497	Não há diferença

* A cor amarela indica que a especificidade apresentou p.valor inferior a 0,05 para análise em questão, porém a análise a ser considerada seria a do tipo oposto a dela.

APÊNDICE 13 - TABELA COMPARATIVA DRB1 MULATO LIHC e LGMH

TABELA 31. Tabeka comparativa DRB1 etnia mulato - LIHC e LGMH

Mulatos		Yates			Pearson		
HLA-							
DRB1	Análise Recomendada	Qui quadrado	Valor de P		Qui Quadrado	Valor de P	
0101	Yates	0,435989051	0,509064751	Não há diferença	1,627725857	0,202017979	Não há diferença
0102	Yates	0,016024978	0,899265069	Não há diferença	0,167266634	0,68255255	Não há diferença
0301	Yates	1,035369984	0,308901042	Não há diferença	1,91354808	0,166569749	Não há diferença
0302	Yates	0,106597725	0,744051331	Não há diferença	1,912119749	0,166728069	Não há diferença
0401	Yates	0,082084552	0,774492117	Não há diferença	1,075102881	0,299795887	Não há diferença
0402	Yates	0,106597725	0,744051331	Não há diferença	1,912119749	0,166728069	Não há diferença
0405	Yates	0,106597725	0,744051331	Não há diferença	1,912119749	0,166728069	Não há diferença
0411	Yates	0,016024978	0,899265069	Não há diferença	0,438458568	0,50786738	Não há diferença
0701	Yates	2,035361842	0,153677658	Não há diferença	3,152046784	0,075831804	Não há diferença
0802	Yates	0,106597725	0,744051331	Não há diferença	0,532619776	0,465507532	Não há diferença
0804	Yates	0,892183535	0,344886282	Não há diferença	2,190775681	0,138839422	Não há diferença
0807	Yates	0,106597725	0,744051331	Não há diferença	0,532619776	0,465507532	Não há diferença
1001	Yates	0,016024978	0,899265069	Não há diferença	0,438458568	0,50786738	Não há diferença
1101	Yates	0,041436439	0,838697804	Não há diferença	0,347652219	0,55544517	Não há diferença
1102	Yates	0,142327139	0,705978395	Não há diferença	0,004123557	0,948799092	Não há diferença
1103	Yates	0,106597725	0,744051331	Não há diferença	0,532619776	0,465507532	Não há diferença
1104	Yates	1,474357673	0,224659558	Não há diferença	3,859649123	0,04946068	Há diferença *
1201	Yates	0,082084552	0,774492117	Não há diferença	1,075102881	0,299795887	Não há diferença
1301	Yates	0,047862712	0,826824945	Não há diferença	0,490114174	0,48387638	Não há diferença
1302	Yates	0,172339952	0,678040536	Não há diferença	0,542763158	0,461289838	Não há diferença
1401	Yates	0,325770618	0,568160179	Não há diferença	0,002003972	0,964294033	Não há diferença
1402	Yates	0,435989051	0,509064751	Não há diferença	1,627725857	0,202017979	Não há diferença

1501	Yates	1,797434891	0,180023007	Não há diferença	2,981649897	0,084213412	Não há diferença
1503	Yates	0,443256579	0,505554818	Não há diferença	1,029239766	0,310337535	Não há diferença
1602	Yates	0,016024978	0,899265069	Não há diferença	0,167266634	0,68255255	Não há diferença
9012	Yates	0,106597725	0,744051331	Não há diferença	0,532619776	0,465507532	Não há diferença

* A cor amarela indica que a especificidade apresentou p.valor inferior a 0,05 para análise em questão, porém a análise a ser considerada seria a do tipo oposto a dela.

